

Worcester Polytechnic Institute Digital WPI

Masters Theses (All Theses, All Years)

Electronic Theses and Dissertations

2004-04-29

Bayesian Predictive Inference Under Informative Sampling and Transformation

Gang Shen

Worcester Polytechnic Institute

Follow this and additional works at: <https://digitalcommons.wpi.edu/etd-theses>

Repository Citation

Shen, Gang, "Bayesian Predictive Inference Under Informative Sampling and Transformation" (2004). *Masters Theses (All Theses, All Years)*. 494.

<https://digitalcommons.wpi.edu/etd-theses/494>

This thesis is brought to you for free and open access by [Digital WPI](#). It has been accepted for inclusion in Masters Theses (All Theses, All Years) by an authorized administrator of Digital WPI. For more information, please contact wpi-etd@wpi.edu.

Bayesian Predictive Inference Under Informative Sampling and Transformation

by

Gang Shen

A Thesis

Submitted to the Faculty

of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Applied Statistics

May 2004

APPROVED:

Balgobin Nandram, Major Professor

Bogdan M. Vernescu, Department Head

Abstract

We have considered the problem in which a biased sample is selected from a finite population, and this finite population itself is a random sample from an infinitely large population, called the superpopulation. The parameters of the superpopulation and the finite population are of interest. There is some information about the selection mechanism in that the selection probabilities are linearly related to the measurements. This is typical of establishment surveys where the selection probabilities are taken to be proportional to the previous year's characteristics. When all the selection probabilities are known, as in our problem, inference about the finite population can be made, but inference about the distribution is not so clear. For continuous measurements, one might assume that the values are normally distributed, but as a practical issue normality can be tenuous. In such a situation a transformation to normality may be useful, but this transformation will destroy the linearity between the selection probabilities and the values. The purpose of this work is to address this issue. In this light we have constructed two models, an ignorable selection model and a nonignorable selection model. We use the Gibbs sampler and the sample importance re-sampling algorithm to fit the nonignorable selection model. We have emphasized estimation of the finite population parameters, although within this framework other quantities can be estimated easily. We have found that our nonignorable selection model can correct the bias due to unequal selection probabilities, and it provides improved precision over the estimates from the ignorable selection model.

In addition, we have described the case in which all the selection probabilities are unknown. This is useful because many agencies (e.g., government) tend to hide these selection probabilities when public-used data are constructed. Also, we have given an extensive theoretical discussion on Poisson sampling, an underlying sampling scheme in our models especially useful in the case in which the selection probabilities are unknown.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my thesis advisor, Professor Balgobin Nandram, for his collaboration on my thesis work. I also wish to thank Dr. Jai Won Choi, National Center for Health Statistics, for providing financial support for me during summer 2003 through Professor Nandram's contract. On this occasion, I also want to extend my sincere appreciation to the statistics faculty at Worcester Polytechnic Institute who helped me so much with my academic life over the past two years. This thesis is dedicated to my beloved wife, Jiong Chen, and son, Jin Jin, to whom I owe much love.

Chapter 1

INTRODUCTION

Inference of quantities, that determine the distribution from which a finite population is generated, can usually be made with a random sample from that finite population. For example, if a random sample is drawn from the finite population, and the sample values come from a normal distribution, it is easy to make inference about the parameters of the distribution or the finite population parameters. However, if individuals in the population have unequal probabilities to be sampled, purposively or not, then a biased inference of the quantities may thus result. This problem is usually referred to as selection bias. The problem is more complex, if the sample values do not follow a simple normal distribution.

In practice, the selection probability π_i for each individual in a finite population often has the following structure,

$$\pi_i = \beta_0 + \beta_1 y_i + e_i, \quad i = 1, \dots, N$$

where y_i , $i = 1, \dots, N$ comprise the finite population from an unknown distribution; β_0, β_1 are unknown regression coefficients; e_i , $i = 1, \dots, N$ are the errors, a random sample from a normal distribution with mean 0 and variance σ_e^2 . If, in addition, one can assume that the y_i come from a normal distribution with mean μ and variance σ^2 , inference can be made about μ and the finite population mean. However, this approach has the problem that the π_i are not random variables (i.e., they are pure numbers specified by the design). Thus, this approach introduces an undesirable source of variation.

An alternative procedure is to take

$$y_i = \beta_0 + \beta_1 \pi_i + e_i, \quad i = 1, \dots, N.$$

Using the sampled y_i , one can easily make inference about the finite population. However,

this approach can not be used to make inference about the parameters of the superpopulation because we can not model the y_i in two different ways. Also, inference about the finite population mean may require a transformation to normality.

In addition to these problems, in sample surveys the y_i are typically not normally distributed. The question then arises, “What procedure can be used to obtain inference for both the finite population mean and the superpopulation mean when there is selection bias and the data are not normally distributed?”

There are two general approaches to inference about the superpopulation parameters when there is informative sampling; i.e., use either an ignorable or a nonignorable model. In the nonignorable model the selection probabilities are related to the responses, and in the ignorable model they are not; see Sugden and Smith (1984), or for a brief discussion, Krieger and Pfeffermann (1992). We have used a nonignorable model. Referring to the ignorable model Pfeffermann, Krieger and Rinott (1998) state that “A different approach in wide use to deal with the effects of informative sampling is to replace the ordinary sample estimates or estimating equations by weighted analogies obtained by weighting the sample observations inversely proportional to the sample selection probabilities. The use of this approach is restricted in general to point estimation and does not permit the use of standard inference tools such as likelihood based inference or residual analysis. Probabilistic statements require large sample normality assumptions.”

As noted in the quotation from Pfeffermann et al. (1998), the ignorable model uses weighted sample quantities or estimating equations to estimate superpopulation parameters. For example, Pfeffermann, Skinner, Holmes, Goldstein and Rasbash (1998) use probability-weighted iterative generalized least squares (PWIGLS) with a scaling technique starting with the “census” likelihood. You and Rao (2002) use a two-step approach to obtain design-consistent small area estimates utilizing survey weights.

For the estimation of a finite population quantity, the problem is more complex than inference for the superpopulation parameters because if there is a bias which tends to make the sampled values “large”, the nonsampled values will tend to be “small” (e.g., Patil and Rao 1978, Krieger and Pfeffermann 1992). Such an adjustment could be carried out using our approach, but is very difficult to do using the other nonignorable methodologies that

have been proposed.

Recent research on inference using samples from finite populations has tackled several difficult problems, including the presence of a selection bias and the availability of only a limited amount of information about the sample design. Chambers, Dorfman and Wang (1998), henceforth CDW, assume that one wishes to model the population process that yields the finite population of survey variables. They assume that the only information about the survey design available to the survey analyst is the set of first-order inclusion probabilities for the sampled units. While CDW provide a brief theoretical framework for their (maximum likelihood) procedure, they note that “it is almost impossible to proceed without fixing ideas on an example.” The example that they use is a generalization of one presented by Krieger and Pfeffermann (1992) whose objective is to investigate inferential methods when there is a selection bias. Thus, CDW analyze a situation where there is limited information about the survey design and there is a selection bias. While the model that CDW examine is of interest because it permits a theoretical investigation of this complicated situation, it is of somewhat limited practical value because it assumes that the finite population is generated as a random sample from a normal distribution.

Nandram and Sedransk (2004) relax this assumption by requiring only a transformation to normality. The purpose of their paper is to demonstrate, by examples, the value of using Bayesian methods in complicated sample survey situations such as this one; i.e., where there is selection bias and limited sample information. While completely general solutions to such problems are not available because of marked differences in the assumptions, their specification should be close to those seen in many surveys. For example, in establishment surveys the selection probability is often proportional to a measure of size which is linearly related to the variable of interest.

Pfeffermann, Krieger and Rinott (1998) consider problems similar to the one investigated in Nandram and Sedransk (2004) in that they assume that the first-order selection probabilities are related to the response variables and these probabilities are known only for the sampled units. To make inference for the superpopulation parameters they derive marginal likelihoods using weighted distributions in the spirit of Patil and Rao (1978). However, to obtain the joint likelihood they have to use asymptotic arguments to justify combining

the marginal likelihoods. Moreover, their methodology permits inference only for the superpopulation parameters. In their framework, extension to inference for finite population parameters is difficult; See Krieger and Pfeiffermann (1992) and Pfeiffermann and Sverchkov (1999) for related work. To incorporate selection bias, Malec, Davis, and Cao (1999) use a hierarchical Bayesian method to estimate a finite population mean when there are binary data. Difficulty in including the selection probabilities directly in the model forces them to make an ad hoc adjustment to the likelihood function and to use a Bayes empirical Bayes (i.e., not a full Bayesian) approach.

Burgos and Nandram (2003) discussed a situation with selection bias where inclusion probabilities of all individuals in the population are known. By introducing exterior variables $\phi_i, i = 1, \dots, N$ into the model and then setting all the ϕ_i to be 0 except ϕ_N , they successfully included the selection bias in the model. In Chapter 2 of this paper, we discuss the same situation by a slight different approach. We also introduce in the exterior variables $\phi_i, i = 1, \dots, N - 1$ to include the selection bias. Using a non one-to-one transformation and some complex algebraic operations, we successfully exclude all the exogenous variables in the model (i.e., our method does not include ϕ_N). Besides, we do not restrict the topic only in the case where the finite population is generated as a random sample from a normal distribution. In chapter 3, we extend our topic to a generalized case where only a transformation to normality is required. In chapter 4, we present a real case where our generalized model discussed in chapter 3 is appropriate and applicable.

For data simulation of the models, we devised two sampling methods in our paper for a situation where individuals in the population are sampled with unequal inclusion probabilities and without replacement. One is called PPS sampling, that is first we calculated the cumulative first-order inclusion probabilities for each individual in the population, noting that sum of these probabilities should be equal to the sample size n . Then we draw a random number u from Uniform(0, 1) and calculate $u + (k - 1), k = 1, \dots, n$; we pick up the individuals whose cumulative first-order inclusion probabilities interval includes a $u + (k - 1)$. Note that because the first order inclusion probability is supposed to be no greater than 1, thus there are no replicates in a sample. The other one is called Poisson sampling. That is for each individual, we draw a random number u from Uniform(0, 1). If u

is less than the individual's inclusion probability, then the individual will be included in the sample, otherwise excluded. This procedure is run for all the individuals, keeping a sample only when the number of individuals included in the sample is exactly the desired sample size n ; otherwise the procedure is restarted for all the individuals in the population. As the expectation of the number of individuals to be included is exactly n , we may expect high efficiency of this sampling method.

The plan of this thesis is as follows. In Chapter 2 we discuss the situation in which the selection probabilities are “proportional” to the population values which are assumed to be normally distributed. In Chapter 3 we discuss the more important and practical situation in which the selection probabilities are “proportional” to the population values which are transformed to satisfy this linearity assumption. Also, in this situation under a further transformation, the transformed population values, are assumed to be normally distributed. We use the Gibbs sampler and the SIR algorithm to perform the computations. In Chapter 4, we present an illustration using an example on natural gas production. In Appendix A, we show how to estimate the selection probabilities when they are all assumed to be unknown, but where our linearity assumption is expected to hold. In Appendix B, we present a theoretical discussion on modified Poisson sampling.

Chapter 2

MODEL WITH ALL INCLUSION PROBABILITIES KNOWN

2.1 Model Assumption

We consider a situation in which there is a selection bias when a sample is drawn from a finite population. Let π_i , $i = 1, \dots, N$ denote the set of selection probabilities, $0 \leq \pi_i \leq 1$ and π_i are all known. Let y_i denote the corresponding response variable, $i = 1, \dots, N$. Then we assume that

$$\pi_i \propto \beta_0 + y_i + e_i, \quad i = 1, \dots, N$$

where e_i are errors. That is, the sample design is informative, and the π_i are “proportional” to the $\beta_0 + y_i$ with noise in the proportionality. Now any sample design must satisfy $\sum_{i=1}^N \pi_i = n$, the sample size. Thus,

$$\pi_i = \frac{n(\beta_0 + y_i + e_i)}{N(\beta_0 + \bar{y} + \bar{e})}, \quad i = 1, \dots, N$$

where $\bar{y} = \frac{\sum_{i=1}^N y_i}{N}$ and $\bar{e} = \frac{\sum_{i=1}^N e_i}{N}$.

Now let $\nu_i = \beta_0 + y_i + e_i$, $i = 1, \dots, N$ where ν_i is a latent variable, we have

$$\pi_i = \frac{n\nu_i}{N\bar{\nu}}, \quad i = 1, \dots, N.$$

Note that because $0 \leq \pi_i \leq 1$, the ν_i must all be non-negative or non-positive. We take $\nu_i \geq 0$, $i = 1, \dots, N$. We also take $c_i = \frac{N}{n}\pi_i$, $i = 1, \dots, N$, thus $\sum_{i=1}^N c_i = N$.

We assume that the response variables

$$y_i | \mu, \sigma^2 \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2), \quad i = 1, \dots, N.$$

This is a standard assumption for a random sample drawn from the population also. However, because of the selection bias, this assumption fails for both the sampled individuals

and the non-sampled individuals. Let \underline{y}_s denote the vector of sampled values, and \underline{y}_{ns} the vector of non-sampled values. Then the vector of all population values is $\underline{y} = (\underline{y}'_s, \underline{y}'_{ns})'$.

2.2 Main Results

Given $\beta_0, \mu, \sigma^2, \sigma_e^2$, (ν_i, y_i) are independent with joint density function, it follows that

$$f(\nu_i, y_i | \beta_0, \mu, \sigma^2, \sigma_e^2) = \frac{\frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{1}{2\sigma_e^2}(\nu_i - \beta_0 - y_i)^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2}}{\Phi\left(\frac{\beta_0 + \mu}{\sqrt{\sigma^2 + \sigma_e^2}}\right)},$$

$$i = 1, \dots, N, \quad \nu_i > 0, \quad -\infty < y_i < \infty$$

where $\Phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$, $-\infty < t < \infty$, the standard normal cumulative distribution function. Thus, the joint density function

$$f(\underline{\nu}, \underline{y} | \beta_0, \mu, \sigma^2, \sigma_e^2) = \frac{\prod_{i=1}^N \left\{ \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{1}{2\sigma_e^2}(\nu_i - \beta_0 - y_i)^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2} \right\}}{\left\{ \Phi\left(\frac{\beta_0 + \mu}{\sqrt{\sigma^2 + \sigma_e^2}}\right) \right\}^N},$$

$$\nu_i > 0, \quad -\infty < y_i < \infty, \quad i = 1, \dots, N.$$

Next we incorporate the restriction that $\pi_i = \frac{n\nu_i}{N\bar{\nu}}$, $i = 1, \dots, N$. We consider the transformed variable $\phi_i = \nu_i - c_i\bar{\nu}$, $i = 1, \dots, N-1$. Note that because the π_i sum to n , we have $N-1$ degrees of freedom. This is different from Burgos and Nandram (2003). So that our transformation is from a N -dimension space to a $(N-1)$ -dimension space. Let $\underline{\phi}'_{(N)} = (\phi_1, \dots, \phi_{N-1})$, $\underline{C}'_{(N)} = (c_1, \dots, c_{N-1})$. Then,

$$\underline{\phi}_{(N)} = B' \underline{\nu}$$

where $B' = \left(I_{N-1} - \frac{\underline{C}_{(N)} \underline{1}'_{N-1}}{N}, \quad -\frac{\underline{C}_{(N)}}{N} \right)$, the $(N-1) \times N$ transformation matrix. It follows that,

$$\begin{aligned} f(\underline{\phi}_{(N)}, \underline{y} | \beta_0, \mu, \sigma^2, \sigma_e^2) &= \frac{1}{|2\pi\sigma_e^2 B' B|^{\frac{1}{2}}} \times \frac{\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2}}{\left\{ \Phi\left(\frac{\beta_0 + \mu}{\sqrt{\sigma^2 + \sigma_e^2}}\right) \right\}^N} \\ &\times e^{-\frac{1}{2}(\underline{\phi}_{(N)} - B'(\underline{y} + \beta_0 \underline{1}))' (B' B \sigma_e^2)^{-1} (\underline{\phi}_{(N)} - B'(\underline{y} + \beta_0 \underline{1}))}. \end{aligned}$$

We incorporate the restrictions by taking $\phi_{(N)} = \mathbf{0}$. That is,

$$f(\phi_{(N)} = \mathbf{0}, \mathbf{y}|\beta_0, \mu, \sigma^2, \sigma_e^2) = \frac{1}{|2\pi\sigma_e^2 B' B|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma_e^2}(\mathbf{y} + \beta_0 \mathbf{1})' B(B' B)^{-1} B'(\mathbf{y} + \beta_0 \mathbf{1})} \\ \times \frac{\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2}}{\left\{ \Phi\left(\frac{\beta_0 + \mu}{\sqrt{\sigma^2 + \sigma_e^2}}\right) \right\}^N}.$$

It is convenient to make the re-parameterization, $\rho = \frac{\sigma^2}{\sigma^2 + \sigma_e^2}$, so that $\sigma_e^2 = \frac{1-\rho}{\rho}\sigma^2$. Thus our new parameters are $\beta_0, \mu, \sigma^2, \rho$ and note that $0 < \rho < 1$. It is interesting that we now have ρ bounded. Before we had $\beta_0, \mu, \sigma^2, \sigma_e^2$ all unbounded. Thus we have

$$f(\phi_{(N)} = \mathbf{0}, \mathbf{y}|\beta_0, \mu, \sigma^2, \rho) = \left(\frac{\rho}{1-\rho}\right)^{\frac{N-1}{2}} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{2N-1}{2}} \frac{1}{|B' B|^{\frac{1}{2}}} \\ \times \frac{e^{-\frac{1}{2\sigma^2} \left\{ \frac{\rho}{1-\rho}(\mathbf{y} + \beta_0 \mathbf{1})' B(B' B)^{-1} B'(\mathbf{y} + \beta_0 \mathbf{1}) + \sum_{i=1}^N (y_i - \mu)^2 \right\}}}{\left\{ \Phi\left(\frac{(\beta_0 + \mu)\sqrt{\rho}}{\sigma}\right) \right\}^N}. \quad (2.1)$$

Note that B is a matrix of constants (i.e., it does not depend on $\mathbf{y}, \beta_0, \mu, \sigma^2, \rho$).

Let I_i denote the selection indicators, $i = 1, \dots, N$, and \mathbf{I} the vector of selection indicators.

Thus,

$$P(\mathbf{I}, \phi_{(N)}, \mathbf{y}|\beta_0, \mu, \sigma^2, \rho) = P(\mathbf{I}|\phi_{(N)}, \mathbf{y}, \beta_0, \mu, \sigma^2, \rho) \times P(\phi_{(N)}, \mathbf{y}|\beta_0, \mu, \sigma^2, \rho)$$

where $P(\mathbf{I}|\phi_{(N)}, \mathbf{y}, \beta_0, \mu, \sigma^2, \rho) = \prod_{i=1}^N \left\{ \pi_i^{I_i} (1 - \pi_i)^{1-I_i} \right\}$ under Poisson sampling. Because the π_i are all known and indicators I_i are all observed, the term $P(\mathbf{I}|\phi_{(N)}, \mathbf{y}, \beta_0, \mu, \sigma^2, \rho)$ is a constant. For a full Bayesian analysis, we take $\beta_0, \mu, \sigma^2, \rho$ independent with

$$P(\beta_0) = 1, \quad -\infty < \beta_0 < \infty,$$

$$P(\mu) = 1, \quad -\infty < \mu < \infty,$$

$$\sigma^{-2} \sim \text{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right), \quad a = .002, \quad b = .002$$

and

$$\rho \sim \text{Uniform}(0, 1).$$

Thus, the joint prior density is

$$\pi(\beta_0, \mu, \sigma^2, \rho) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}}, \quad -\infty < \beta_0, \mu < \infty, \quad 0 < \rho < 1, \quad \sigma^2 > 0.$$

Note that this is a proper prior in σ^2 and ρ but not in β_0, μ . Thus, by Bayes' theorem the joint posterior density of $(y_{ns}, \beta_0, \mu, \sigma^2, \rho)$ is

$$\begin{aligned} \pi(\phi_{(N)}, y_{ns}, \beta_0, \mu, \sigma^2, \rho | I, y_s) &\propto P(I, \phi_{(N)}, y, \beta_0, \mu, \sigma^2, \rho) \\ &= P(I, \phi_{(N)}, y | \beta_0, \mu, \sigma^2, \rho) \pi(\beta_0, \mu, \sigma^2, \rho). \end{aligned}$$

Now, incorporating the constraints we have

$$\pi(\phi_{(N)} = 0, y_{ns}, \beta_0, \mu, \sigma^2, \rho | I, y_s) \propto P(I, \phi_{(N)} = 0, y | \beta_0, \mu, \sigma^2, \rho) \pi(\beta_0, \mu, \sigma^2, \rho).$$

Thus, posterior inference about $y_{ns}, \beta_0, \mu, \sigma^2, \rho$ is based on $\pi(\phi_{(N)} = 0, y_{ns}, \beta_0, \mu, \sigma^2, \rho | I, y_s)$, where

$$\begin{aligned} \pi(\phi_{(N)} = 0, y, \beta_0, \mu, \sigma^2, \rho | I, y_s) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}} \left(\frac{\rho}{1-\rho}\right)^{\frac{N-1}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{2N-1}{2}} \\ &\times \frac{e^{-\frac{1}{2\sigma^2} \left\{ \frac{\rho}{1-\rho} (y + \beta_0 \mathbf{1})' B (B' B)^{-1} B' (y + \beta_0 \mathbf{1}) + \sum_{i=1}^N (y_i - \mu)^2 \right\}}}{\left\{ \Phi \left(\frac{(\beta_0 + \mu) \sqrt{\rho}}{\sigma} \right) \right\}^N}. \end{aligned}$$

Because the joint posterior density function is complex, we use Markov chain Monte Carlo method to draw a “random” sample from it. We will show how to use the Gibbs sampler and the sample importance re-sampling (SIR) algorithm to get the sample.

It is convenient to transform ρ to $\tau = \frac{\rho}{1-\rho}$ keeping $y_{ns}, \beta_0, \mu, \sigma^2$ untransformed. Thus,

$$\begin{aligned} \pi(\phi_{(N)} = 0, y_{ns}, \beta_0, \mu, \sigma^2, \tau | I, y_s) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}} \frac{1}{(1+\tau)^2} \tau^{\frac{N-1}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{2N-1}{2}} \\ &\times \frac{e^{-\frac{1}{2\sigma^2} \left\{ \tau (y + \beta_0 \mathbf{1})' B (B' B)^{-1} B' (y + \beta_0 \mathbf{1}) + \sum_{i=1}^N (y_i - \mu)^2 \right\}}}{\left\{ \Phi \left(\frac{(\beta_0 + \mu) \sqrt{\rho}}{\sigma} \right) \right\}^N}, \end{aligned}$$

where $\frac{1}{(1+\tau)^2}$ is the Jacobian of the transformation.

2.3 Computation

To perform the computation, we write

$$\pi(\phi_{(N)} = 0, y_{ns}, \beta_0, \mu, \sigma^2, \tau | I, y_s) = KR(\beta_0, \mu, \sigma, \tau) \pi_a(y_{ns}, \beta_0, \mu, \sigma^2, \tau | I, y_s) \quad (2.2)$$

where K is the normalization constant,

$$R(\beta_0, \mu, \sigma, \tau) = \left\{ (1 + \tau)^2 \left[\Phi \left(\frac{(\beta_0 + \mu)}{\sigma} \sqrt{\frac{\tau}{1 + \tau}} \right) \right]^N \right\}^{-1}$$

and

$$\begin{aligned} \pi_a(y_{ns}, \beta_0, \mu, \sigma^2, \tau | I, y_s) &\propto \tau^{\frac{N-1}{2}} \left(\frac{1}{\sigma^2} \right)^{\frac{2N+a+1}{2}} \\ &\times e^{-\frac{1}{2\sigma^2} \left\{ \tau(y + \beta_0 \mathbf{1})' B(B'B)^{-1} B' (y + \beta_0 \mathbf{1}) + \sum_{i=1}^N (y_i - \mu)^2 + b \right\}}. \end{aligned} \quad (2.3)$$

Thus, we may use the Gibbs sampler to draw a sample from $\pi_a(y_{ns}, \beta_0, \mu, \sigma^2, \tau | I, y_s)$ and the SIR algorithm to “convert” this sample to the one from $\pi(y_{ns}, \beta_0, \mu, \sigma^2, \tau | I, y_s)$. To perform the Gibbs sampler under $\pi_a(y_{ns}, \beta_0, \mu, \sigma^2, \tau | I, y_s)$, we need the conditional posterior density of each parameter given all the others. First we note that

$$y | \beta_0, \mu, \sigma^2, \tau \sim \text{Normal} \left\{ [I + \tau B(B'B)^{-1} B']^{-1} (\beta_0 + \mu) \mathbf{1} - \beta_0 \mathbf{1}, \quad \sigma^2 [I + \tau B(B'B)^{-1} B']^{-1} \right\}$$

and

$$(I + \tau B(B'B)^{-1} B')^{-1} = I - \rho c(c'c)^{-1} c'.$$

So

$$y | \beta_0, \mu, \sigma^2, \tau \sim \text{Normal} \left\{ [I - \rho B(B'B)^{-1} B'] (\beta_0 + \mu) \mathbf{1} - \beta_0 \mathbf{1}, \quad \sigma^2 [I - \rho B(B'B)^{-1} B'] \right\}.$$

Further, we note that

$$B(B'B)^{-1} B' = I - c(c'c)^{-1} c',$$

where $c' = (c_1, \dots, c_N)$ and I is the $N \times N$ identity matrix; see Appendix B. So,

$$y | \beta_0, \mu, \sigma^2, \tau \sim \text{Normal} \left\{ \left[(1 - \rho)I + \rho \frac{cc'}{c'c} \right] (\beta_0 + \mu) \mathbf{1} - \beta_0 \mathbf{1}, \quad \sigma^2 \left[(1 - \rho)I + \rho \frac{cc'}{c'c} \right] \right\}.$$

Hence,

$$y_{ns}|y_s, \beta_0, \mu, \sigma^2, \tau \sim \text{Normal} \left\{ \mu \underline{1}_{ns} + \frac{c_s'(y_s - \mu \underline{1}_s)}{\tau^{-1} \underline{c}' \underline{c} + \underline{c}'_s \underline{c}_s} \underline{c}_{ns} \right. \\ \left. + \frac{(\beta_0 + \mu) \tau}{1 + \tau} \left[\frac{\tau^{-1} \underline{c}' \underline{1} + \underline{c}'_s \underline{1}_s}{\tau^{-1} \underline{c}' \underline{c} + \underline{c}'_s \underline{c}_s} \underline{c}_{ns} - \underline{1}_{ns} \right], \right. \\ \left. \frac{\sigma^2}{1 + \tau} \left[I + \frac{\underline{c}_{ns} \underline{c}'_{ns}}{\tau^{-1} \underline{c}' \underline{c} + \underline{c}'_s \underline{c}_s} \right] \right\}.$$

It is easy to show that

$$y_{ns}|y_s, \beta_0, \mu, \sigma^2, \tau \sim \text{Normal} \left\{ \left(\frac{\mu}{1 + \tau} - \beta_0 \right) \underline{1}_{ns} + [\lambda g(\mu) + (1 - \lambda) \bar{y}_{ws}] \underline{C}_{ns}, \right. \\ \left. \frac{\sigma^2}{1 + \tau} \left[I + \frac{\tau \lambda}{\sum_{i=1}^N c_i^2} \underline{C}_{ns} \underline{C}'_{ns} \right] \right\}$$

where $\lambda = \frac{\sum_{i=1}^N c_i^2}{\sum_{i=1}^N c_i^2 + \tau \sum_{i \in s} c_i^2}$, $g(\mu) = \frac{\tau(N - \sum_{i \in s} c_i)}{(1 + \tau) \sum_{i=1}^N c_i^2} \mu$ and $\bar{y}_{ws} = \frac{\sum_{i \in s} c_i(\beta_0 + y_i)}{\sum_{i \in s} c_i^2}$. This form is useful because one can deduce various scenarios about selection bias. For example, if σ_e^2 is small, there is a tight linear relation between the π_i and the y_i . Then, if the y_i are very variable, there is selection bias, and the model can adjust for it.

The conditional posterior densities of μ, σ^2, τ under $\pi_a(y_{ns}, \mu, \sigma^2, \tau | \underline{I}, y_s)$ are

$$\beta_0|y, \mu, \sigma^2, \tau \sim \text{Normal} \left(\frac{\bar{y}_w - \bar{y}}{1 - \frac{N}{C'C}}, \frac{\sigma^2}{\tau N(1 - \frac{N}{C'C})} \right), \\ \mu|y, \beta_0, \sigma^2, \tau \sim \text{Normal} \left(\bar{y}, \frac{\sigma^2}{N} \right), \\ \sigma^{-2}|y, \beta_0, \mu, \tau \sim \text{Gamma} \left(\frac{2N + a - 1}{2}, \frac{\tau A + H + b}{2} \right), \\ \tau|y, \beta_0, \mu, \sigma^2 \sim \text{Gamma} \left(\frac{N + 1}{2}, \frac{A}{2\sigma^2} \right),$$

where $\bar{y}_w = \frac{\sum_{i=1}^N c_i y_i}{\sum_{i=1}^N c_i^2}$, $\bar{y} = \frac{\sum_{i=1}^N y_i}{N}$, $A = \sum_{i=1}^N (\beta_0 + y_i)^2 - \frac{[\sum_{i=1}^N c_i(\beta_0 + y_i)]^2}{\sum_{i=1}^N c_i^2}$ and $H = \sum_{i=1}^N (y_i - \mu)^2$.

Finally, we note that we can draw $y_{ns} | \beta_0, \mu, \sigma^2, \tau, y_s$ using the product rule. It is easy

to show that

$$y_{n+k+1} | y_{n+k}, \beta_0, \mu, \sigma^2, \rho \sim \text{Normal} \left\{ \mu + \frac{c_s'(y_s - \mu \underline{1}_s)}{\tau^{-1} \underline{c}' \underline{c} + \underline{c}'_s \underline{c}_s} c_{n+k+1} + \frac{(\beta_0 + \mu) \tau}{1 + \tau} \left[\frac{\tau^{-1} \underline{c}' \underline{1} + \underline{c}'_s \underline{1}_s}{\tau^{-1} \underline{c}' \underline{c} + \underline{c}'_s \underline{c}_s} c_{n+k+1} - 1 \right], \frac{\sigma^2}{1 + \tau} \left[1 + \frac{c_{n+k+1}^2}{\tau^{-1} \underline{c}' \underline{c} + \underline{c}'_s \underline{c}_s} \right] \right\}.$$

Thus by the product rule,

$$f(y_{ns} | y_s, \beta_0, \mu, \sigma^2, \tau) = \prod_{k=n+1}^N f(y_k | y_{(k)}, \beta_0, \mu, \sigma^2, \tau)$$

where $y'_{(k)} = (y_1, \dots, y_{k-1})$, $k = n+1, \dots, N$.

The Gibbs sampler provides a sample $\Omega^{(h)}$, $h = 1, \dots, M$ from the joint posterior density, where $\Omega^{(h)} = (y_{ns}, \beta_0, \mu, \sigma^2, \tau)$. We perform the SIR algorithm by sub-sampling the $\Omega^{(h)}$ with weights

$$W_h = \frac{R(\Omega^{(h)})}{\sum_{h=1}^M R(\Omega^{(h)})}, \quad h = 1, \dots, M.$$

Then, we draw a sample from the discrete probability mass function $\{(\Omega^{(h)}, W_h), h = 1, \dots, M\}$ with replacement.

2.4 Data Simulation

For data simulation of a case where big selection bias exists, we follow the procedure as follows:

- (1) Set $\beta_0 = 0$, $\mu = 2$, $\sigma^2 = 0.2375$, $\sigma_e^2 = 0.0125$, then $\rho = .95$
- (2) Generate y from $\text{Normal}(\mu, \sigma^2)$.
- (3) Generate ν from $\text{Normal}(\beta_0 + y, \sigma_e^2)$ using one-one sampling to get $\nu > 0$.

- (4) Repeat step (1) – (4) until we have 100 y and ν .
- (5) Calculate selection probabilities π_i , $i = 1, \dots, 100$ for each y_i using $\pi_i = \frac{\nu_i}{100\bar{\nu}}$.
- (6) Draw a sample of 25 y without replacement from the population of 100 y using Poisson sampling.
- (7) Pretending we don't know the 75 non-sampled y , draw a sample of size 75 from the joint posterior distribution as obtained in the model.
- (8) Make inference about quantities we are interested in, like μ and the finite population mean, from the sample we got in step(8)
- (9) Repeat step(2) – (9) for 1000 times.

Now we may compare these inferences of the quantities with the “true” values.

Table1: Comparison of the ignorable model (IG) and the nonignorable model (NIG) in inference of the parameter μ in the case where big the selection bias exists:

IG	NIG	RAVG	RSTD	Coverage by 95% Interval	
				IG	NIG
1.058	0.992	0.534	0.070	0.803	0.943

Table2: Comparison of the ignorable model (IG) and the nonignorable model (NIG) in inference of the finite population mean \bar{y} in the case where big the selection bias exists:

IG	NIG	RAVG	RSTD	Coverage by 95% Interval	
				IG	NIG
1.058	0.992	0.225	0.051	0.742	0.855

From the table, one can easily see that in the situation where big selection bias exists, average estimates of μ and \bar{y} by the ignorable model (IG) is 5.8% away from the true values, while that by nonignorable is only 0.8% away from the “true” value; The average ratio of standard deviation (RAVG) of the estimate of μ by nonignorable model to that by ignorable model is only 0.534, which means the precision of the estimate of μ by nonignorable model is almost as twice as the ignorable model. More precision is gained by the nonignorable model in estimate of population mean compared with the ignorable model, (0.225). The standard deviation of this average ratio (RSTD) is very small, say 0.070 for μ and 0.051 for \bar{y} . As for the 95% creditable interval coverage, nonignorable is much better than ignorable model for both estimates of μ and \bar{y} , say 0.943 vs 0.803 and .855 vs 0.742.

We may also reset the parameters $\beta_0 = 0$, $\mu = 2$, $\sigma^2 = 0.0095$, $\sigma_e^2 = 0.0005$ and repeat the procedures described as above for data simulation of a case where small selection bias exists. Here is the result.

Table3: Comparison of the ignorable model (IG) and the nonignorable model (NIG) in inference of the parameter μ in the case where small the selection bias exists:

IG	NIG	RAVG	RSTD	<u>Coverage by 95% Interval</u>	
				IG	NIG
1.0026	0.9996	0.5327	0.069	0.938	0.949

Table4: Comparison of the ignorable model (IG) and the nonignorable model (NIG) in inference of the finite population mean \bar{y} in the case where small the selection bias exists:

IG	NIG	RAVG	RSTD	<u>Coverage by 95% Interval</u>	
				IG	NIG
1.0026	0.9995	0.232	0.050	0.933	0.941

From the table above, one can easily see that in the situation where only small selection bias exists, ignorable model works almost as well as our nonignorable model, except that nonignorable model attains much precision in the inference.

2.5 Appendix

2.5.1 A: Ignorable Model

The ignorable selection bias model is

$$y_i|\mu, \sigma \sim N(\mu, \sigma^2), \quad i = 1, 2, \dots, N$$

with prior

$$P(\mu, \sigma^2) \propto \frac{1}{\sigma^2}.$$

It can be shown that the $100(1 - \alpha)\%$ highest posterior density(HPD) intervals for μ and \bar{y} are

$$\bar{y}_s \pm \frac{s}{\sqrt{n}} \times t_{n-1, 1-\alpha/2}$$

and

$$\bar{y}_s \pm \sqrt{1 - f} \times \frac{s}{\sqrt{n}} \times t_{n-1, 1-\alpha/2}$$

respectively, where \bar{y}_s is the sample mean, s is the sample standard deviation, $f = \frac{n}{N}$ is the sample fraction, and $t_{n-1, 1-\alpha/2}$ is the $100(1 - \alpha/2)\%$ percentile of t distribution with $n - 1$ degrees of freedom.

2.5.2 B: Related Facts of Matrix Operation

We present some quantities needed to evaluate the conditional posterior density of y_{ns} .

First we note that

$$B'B = I_{N-1} - \frac{J_{N-1}}{N} + \frac{(1_{N-1} - \underline{c}_{(N)})(1_{N-1} - \underline{c}_{(N)})'}{N},$$

so

$$(B'B)^{-1} = I_{N-1} + J_{N-1} - \frac{(c_N \mathbf{1}_{N-1} - \underline{c}_{(N)})(c_N \mathbf{1}_{N-1} - \underline{c}_{(N)})'}{c'c}$$

where, I_{N-1} is the $(N-1) \times (N-1)$ identity matrix, J_{N-1} is the $(N-1) \times (N-1)$ matrix of ones, $\mathbf{1}_{N-1}$ is the $(N-1) \times 1$ vector of ones, and $c'_{(N)} = (c_1, \dots, c_{N-1})$. It follows that

$$B(B'B)^{-1}B' = I_N - c(c'c)^{-1}c'$$

where $c' = (c_1, \dots, c_N)$, and I_N is the $N \times N$ identity matrix. Also,

$$|B'B| = \frac{1}{N} \left| I_{N-1} - \frac{J_{N-1}}{N} \right| \left\{ N + (\mathbf{1} - \underline{c}_{(N)})' \left(I_{N-1} - \frac{J_{N-1}}{N} \right)^{-1} (\mathbf{1} - \underline{c}_{(N)}) \right\}$$

But because $\left| I_{N-1} - \frac{J_{N-1}}{N} \right| = \frac{1}{N}$ and $\left(I_{N-1} - \frac{J_{N-1}}{N} \right)^{-1} = I_{N-1} + J_{N-1}$, so

$$|B'B| = \frac{c'c}{N^2}.$$

Chapter 3

GENERALIZED MODEL WITH ALL INCLUSION PROBABILITIES KNOWN

We call our model with transformation of the measurements a generalized model.

3.1 Model Assumption

We consider a situation in which there is a selection bias when a sample is drawn from a finite population. Let π_i , $i = 1, \dots, N$ denote the set of selection probabilities, $0 \leq \pi_i \leq 1$ and π_i are all known. Let y_i denote the corresponding response variable, $i = 1, \dots, N$. Then we assume that

$$\pi_i \propto \beta_0 + y_i + e_i, \quad i = 1, \dots, N$$

where e_i are errors. That is, the sample design is informative, and the π_i are “proportional” to the $\beta_0 + y_i$ with noise in the proportionality. Now any sample design must satisfy $\sum_{i=1}^N \pi_i = n$, the sample size. Thus,

$$\pi_i = \frac{n(\beta_0 + y_i + e_i)}{N(\beta_0 + \bar{y} + \bar{e})}, \quad i = 1, \dots, N$$

where $\bar{y} = \frac{\sum_{i=1}^N y_i}{N}$ and $\bar{e} = \frac{\sum_{i=1}^N e_i}{N}$.

Now let $\nu_i = \beta_0 + y_i + e_i$, $i = 1, \dots, N$ where ν_i is a latent variable, we have

$$\pi_i = \frac{n\nu_i}{N\bar{\nu}}, \quad i = 1, \dots, N.$$

Note that because $0 \leq \pi_i \leq 1$, the ν_i must all be non-negative or non-positive. We take $\nu_i \geq 0$, $i = 1, \dots, N$. We also take $c_i = \frac{N}{n}\pi_i$, $i = 1, \dots, N$, thus $\sum_{i=1}^N c_i = N$.

In practice, we may need to make transformation to bring the response variables to a normal distribution. Here, we assume that

$$g(y_i)|\theta_i, \sigma^2 \stackrel{ind}{\sim} \text{Normal}(\theta_i, \sigma^2), \quad i = 1, \dots, N.$$

This is a standard assumption for a random sample drawn from the population also. However, because of the selection bias, this assumption fails for both the sampled individuals and the non-sampled individuals. Let y_s denote the vector of sampled values, and y_{ns} the vector of non-sampled values. Then the vector of all population values is $y = (y'_s, y'_{ns})'$.

Note that the model implies $\nu_i \propto c_i$, so there must exist a constant k such that $\nu_i = kc_i$ for all $i = 1, \dots, N$. Also note that the model assumes $\nu_i = \beta_0 + y_i + e_i$. Thus, intuitively, one may think about using regression method to figure out non-sampled y_i , $i \notin s$. That is, with

$$y_i = -\beta_0 + k_1 c_i + e_i, \quad i \in s$$

we may determine least square estimates of $-\beta_0$ and k_1 by regressing y_i over c_i for all $i \in s$. Then we may use the obtained estimates of $-\beta_0, k_1$ and the given c_i , $i \notin s$ to determine y_i , for all $i \notin s$. Though, regression method can give a point estimate a_i to each y_i , $i = 1, \dots, N$, it doesn't take into consideration the distribution of y . However, we may borrow the strength from regression method. We can use these point estimates a_i as an approximation to the corresponding y_i , $i = 1, \dots, N$. Specifically, we will use a first order Taylor's series expansion on $g(y_i)$ at $y_i = a_i$ for all $i = 1, \dots, N$ in our model as follows.

3.2 Main Results

Given $\beta_0, \theta_i, \sigma^2, \sigma_e^2$, (ν_i, y_i) are independent with joint density function, it follows that

$$f(\nu_i, y_i | \beta_0, \theta_i, \sigma^2, \sigma_e^2) = \frac{\frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{1}{2\sigma_e^2}(\nu_i - \beta_0 - y_i)^2} \times \frac{|g'(y_i)|}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}[g(y_i) - \theta_i]^2}}{K(\beta_0, \theta_i, \sigma^2, \sigma_e^2)},$$

$$i = 1, \dots, N, \quad \nu_i > 0, \quad -\infty < y_i < \infty$$

where

$$K(\beta_0, \theta_i, \sigma^2, \sigma_e^2) = \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{1}{2\sigma_e^2}(\nu_i - \beta_0 - y_i)^2} \times \frac{|g'(y_i)|}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}[g(y_i) - \theta_i]^2} dy_i d\nu_i.$$

Let $g(y_i) = g(a_i) + g'(a_i)(y_i - a_i) + \Delta_i$, where a_i is an approximation of y_i by regression method and $\Delta_i = g(y_i) - g(a_i) - g'(a_i)(y_i - a_i) = o(y_i - a_i)$, then

$$\begin{aligned} f(\nu_i, y_i | \beta_0, \theta_i, \sigma^2, \sigma_e^2) &= \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{1}{2\sigma_e^2}(\nu_i - \beta_0 - y_i)^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(g'(a_i))^2}{2\sigma^2} \left[y_i - \left(a_i + \frac{\theta_i - g(a_i)}{g'(a_i)} \right) \right]^2} \\ &\times \frac{|g'(y_i)| e^{-\frac{\Delta_i}{2\sigma^2} \{ \Delta_i + 2[g(a_i) - \theta_i + g'(a_i)(y_i - a_i)] \}}}{K(\beta_0, \theta_i, \sigma^2, \sigma_e^2)}. \end{aligned}$$

Thus, the joint density function

$$\begin{aligned} f(\nu, \underline{y} | \beta_0, \theta, \sigma^2, \sigma_e^2) &= \left\{ \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-\frac{1}{2\sigma_e^2}(\nu_i - \beta_0 - y_i)^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(g'(a_i))^2}{2\sigma^2} \left[y_i - \left(a_i + \frac{\theta_i - g(a_i)}{g'(a_i)} \right) \right]^2} \right\} \\ &\times \prod_{i=1}^N \frac{|g'(y_i)| e^{-\frac{\Delta_i}{2\sigma^2} \{ \Delta_i + 2[g(a_i) - \theta_i + g'(a_i)(y_i - a_i)] \}}}{K(\beta_0, \theta_i, \sigma^2, \sigma_e^2)}, \\ \nu_i &> 0, \quad -\infty < y_i < \infty, \quad i = 1, \dots, N. \end{aligned}$$

Next we incorporate the restriction that $\pi_i = \frac{n\nu_i}{N\bar{\nu}}$, $i = 1, \dots, N$. We consider the transformed variable $\phi_i = \nu_i - c_i\bar{\nu}$, $i = 1, \dots, N-1$. Note that because the π_i sum to n , we have $N-1$ degrees of freedom. So that our transformation is from an N -dimension space to an $(N-1)$ -dimension space. Let $\phi'_{(N)} = (\phi_1, \dots, \phi_{N-1})$, $\underline{c}'_{(N)} = (c_1, \dots, c_{N-1})$. Then,

$$\phi_{(N)} = B' \underline{\nu}$$

where $B' = \left(I_{N-1} - \frac{\underline{c}_{(N)} \underline{c}'_{(N)}}{N}, \quad -\frac{\underline{c}_{(N)}}{N} \right)$, the $(N-1) \times N$ transformation matrix. It follows that,

$$\begin{aligned} f(\phi_{(N)}, \underline{y} | \beta_0, \theta, \sigma^2, \sigma_e^2) &= \frac{1}{|2\pi\sigma_e^2 B' B|^{\frac{1}{2}}} \times \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{N}{2}} e^{-\frac{1}{2\sigma^2} (\underline{y} - \underline{e})' D (\underline{y} - \underline{e})} \\ &\times e^{-\frac{1}{2} (\phi_{(N)} - B'(\underline{y} + \beta_0 \underline{1}))' (B' B \sigma_e^2)^{-1} (\phi_{(N)} - B'(\underline{y} + \beta_0 \underline{1}))} \\ &\times \prod_{i=1}^N \frac{|g'(y_i)| e^{-\frac{\Delta_i}{2\sigma^2} \{ \Delta_i + 2[g(a_i) - \theta_i + g'(a_i)(y_i - a_i)] \}}}{(K(\beta_0, \theta_i, \sigma^2, \sigma_e^2))}, \end{aligned}$$

where, D is a diagonal matrix, $D = \text{diag}[(g'(a_1))^2, \dots, (g'(a_N))^2]$ and e is column vector, $e = (a_1 + \frac{\theta_1 - g(a_1)}{g'(a_1)}, \dots, a_N + \frac{\theta_N - g(a_N)}{g'(a_N)})'$.

We incorporate the restrictions by taking $\phi_{(N)} = \mathbf{0}$. That is,

$$\begin{aligned} f(\phi_{(N)} = \mathbf{0}, \mathbf{y}|\beta_0, \theta, \sigma^2, \sigma_e^2) &= \frac{1}{|2\pi\sigma_e^2 B' B|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma_e^2}(\mathbf{y} + \beta_0 \mathbf{1})' B (B' B)^{-1} B' (\mathbf{y} + \beta_0 \mathbf{1})} \\ &\times \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{N}{2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{e})' D (\mathbf{y} - \mathbf{e})} \\ &\times \prod_{i=1}^N \frac{|g'(y_i)| e^{-\frac{\Delta_i}{2\sigma^2} \{\Delta_i + 2[g(a_i) - \theta_i + g'(a_i)(y_i - a_i)]\}}}{(K(\beta_0, \theta_i, \sigma^2, \sigma_e^2))}. \end{aligned}$$

It is convenient to make the re-parameterization, $\rho = \frac{\sigma^2}{\sigma^2 + \sigma_e^2}$, so that $\sigma_e^2 = \frac{1-\rho}{\rho}\sigma^2$. Thus our new parameters are $\beta_0, \theta, \sigma^2, \rho$ and note that $0 < \rho < 1$. It is interesting that we now have ρ bounded. Before we had $\beta_0, \theta, \sigma^2, \sigma_e^2$ all unbounded. Thus we have

$$\begin{aligned} f(\phi_{(N)} = \mathbf{0}, \mathbf{y}|\beta_0, \theta, \sigma^2, \rho) &= \left(\frac{\rho}{1-\rho} \right)^{\frac{N-1}{2}} \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{2N-1}{2}} \frac{1}{|B' B|^{\frac{1}{2}}} \\ &\times \prod_{i=1}^N \frac{|g'(y_i)| \exp \left\{ -\frac{\Delta_i}{2\sigma^2} \{\Delta_i + 2[g(a_i) - \theta_i + g'(a_i)(y_i - a_i)]\} \right\}}{(K(\beta_0, \theta_i, \sigma^2, (1-\rho)\sigma^2/\rho))} \\ &\times \exp \left\{ -\frac{1}{2\sigma^2} \left[\frac{\rho}{1-\rho} (\mathbf{y} + \beta_0 \mathbf{1})' B (B' B)^{-1} B' (\mathbf{y} + \beta_0 \mathbf{1}) \right. \right. \\ &\quad \left. \left. + (\mathbf{y} - \mathbf{e})' D (\mathbf{y} - \mathbf{e}) \right] \right\}. \end{aligned}$$

Note that B is a matrix of constants (i.e., it does not depend on $\mathbf{y}, \beta_0, \theta, \sigma^2, \rho$).

Let I_i denote the selection indicators, $i = 1, \dots, N$, and \mathbf{I} is the vector of selection indicators.

Thus,

$$P(\mathbf{I}, \phi_{(N)}, \mathbf{y}|\beta_0, \theta, \sigma^2, \rho) = P(\mathbf{I}|\phi_{(N)}, \mathbf{y}, \beta_0, \theta, \sigma^2, \rho) \times P(\phi_{(N)}, \mathbf{y}|\beta_0, \theta, \sigma^2, \rho)$$

where $P(\mathbf{I}|\phi_{(N)}, \mathbf{y}, \beta_0, \theta, \sigma^2, \rho) = \prod_{i=1}^N \left\{ \pi_i^{I_i} (1 - \pi_i)^{1-I_i} \right\}$ under Poisson sampling. Because the π_i are all known and indicators I_i are all observed, the term $P(\mathbf{I}|\phi_{(N)}, \mathbf{y}, \beta_0, \theta, \sigma^2, \rho)$ is a constant. For a full Bayesian analysis, we take $\beta_0, \theta_i, \sigma^2, \rho$ independent with

$$P(\beta_0) = 1, \quad -\infty < \beta_0 < \infty,$$

$$P(\theta_i) = 1, \quad -\infty < \theta_i < \infty,$$

$$\sigma^{-2} \sim \text{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right), \quad a = .002, \quad b = .002$$

and

$$\rho \sim \text{Uniform}(0, 1).$$

Thus, the joint prior density is

$$\pi(\beta_0, \theta, \sigma^2, \rho) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}}, \quad -\infty < \beta_0, \theta_i < \infty, \quad 0 < \rho < 1, \quad \sigma^2 > 0$$

Note that this is a proper prior in σ^2 and ρ but not in β_0, θ . Thus, by Bayes' theorem the joint posterior density of $(y_{ns}, \beta_0, \theta, \sigma^2, \rho)$ is

$$\begin{aligned} \pi(\phi_{(N)}, y_{ns}, \beta_0, \theta, \sigma^2, \rho | I, y_s) &\propto P(I, \phi_{(N)}, y, \beta_0, \theta, \sigma^2, \rho) \\ &= P(I, \phi_{(N)}, y | \beta_0, \theta, \sigma^2, \rho) \pi(\beta_0, \theta, \sigma^2, \rho). \end{aligned}$$

Now, incorporating the constraints we have

$$\pi(\phi_{(N)} = 0, y_{ns}, \beta_0, \theta, \sigma^2, \rho | I, y_s) \propto P(I, \phi_{(N)} = 0, y | \beta_0, \theta, \sigma^2, \rho) \pi(\beta_0, \theta, \sigma^2, \rho).$$

Thus, posterior inference about $y_{ns}, \beta_0, \theta, \sigma^2, \rho$ is based on $\pi(\phi_{(N)} = 0, y_{ns}, \beta_0, \theta, \sigma^2, \rho | I, y_s)$, where

$$\begin{aligned} \pi(\phi_{(N)} = 0, y, \beta_0, \theta, \sigma^2, \rho | I, y_s) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}} \left(\frac{\rho}{1-\rho}\right)^{\frac{N-1}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{2N-1}{2}} \\ &\times \prod_{i=1}^N \frac{|g'(y_i)| \exp\left\{-\frac{\Delta_i}{2\sigma^2} \{\Delta_i + 2[g(a_i) - \theta_i + g'(a_i)(y_i - a_i)]\}\right\}}{(K(\beta_0, \theta_i, \sigma^2, (1-\rho)\sigma^2/\rho))} \\ &\times \exp\left\{-\frac{1}{2\sigma^2} \left[\frac{\rho}{1-\rho} (y + \beta_0 \mathbf{1})' B (B' B)^{-1} B' (y + \beta_0 \mathbf{1}) \right. \right. \\ &\quad \left. \left. + (y - e)' D (y - e) \right] \right\}. \end{aligned}$$

Because the joint posterior density function is complex, we use Markov chain Monte Carlo method to draw a “random” sample from it. We will show how to use the Gibbs sampler and the sample importance re-sampling (SIR) algorithm to get the sample.

It is convenient to transform ρ to $\tau = \frac{\rho}{1-\rho}$ keeping $\underline{y}_{ns}, \beta_0, \theta, \sigma^2$ untransformed. Thus,

$$\begin{aligned} \pi(\phi_{(N)} = \underline{0}, \underline{y}_{ns}, \beta_0, \theta, \sigma^2, \tau | \underline{I}, \underline{y}_s) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}} \frac{1}{(1+\tau)^2} \tau^{\frac{N-1}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{2N-1}{2}} \\ &\times \prod_{i=1}^N \frac{|g'(y_i)| \exp\left\{-\frac{\Delta_i}{2\sigma^2} \{\Delta_i + 2[g(a_i) - \theta_i + g'(a_i)(y_i - a_i)]\}\right\}}{(K(\beta_0, \theta_i, \sigma^2, \sigma^2/\tau))} \\ &\times \exp\left\{-\frac{1}{2\sigma^2} \left[\tau(\underline{y}' + \beta_0 \underline{1}') B(B'B)^{-1} B'(\underline{y} + \beta_0 \underline{1}) \right. \right. \\ &\quad \left. \left. + (\underline{y}' - \underline{e}') D(\underline{y} - \underline{e}) \right] \right\} \end{aligned}$$

where $\frac{1}{(1+\tau)^2}$ is the Jacobian of the transformation.

3.3 Computation

To perform the computation, we write

$$\pi(\phi_{(N)} = \underline{0}, \underline{y}_{ns}, \beta_0, \theta, \sigma^2, \tau | \underline{I}, \underline{y}_s) = KR(\beta_0, \theta, \sigma, \tau) \pi_a(\underline{y}_{ns}, \beta_0, \theta, \sigma^2, \tau | \underline{I}, \underline{y}_s)$$

where K is the normalization constant,

$$R(\beta_0, \theta, \sigma^2, \tau) = \frac{1}{(1+\tau)^2} \prod_{i=1}^N \frac{|g'(y_i)| \exp\left\{-\frac{\Delta_i}{2\sigma^2} \{\Delta_i + 2[g(a_i) - \theta_i + g'(a_i)(y_i - a_i)]\}\right\}}{(K(\beta_0, \theta_i, \sigma^2, \sigma^2/\tau))}$$

and

$$\begin{aligned} \pi_a(\underline{y}_{ns}, \beta_0, \theta, \sigma^2, \tau | \underline{I}, \underline{y}_s) &\propto \tau^{\frac{N-1}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{2N+a+1}{2}} \\ &\times \exp\left\{-\frac{1}{2\sigma^2} \left[\tau(\underline{y}' + \beta_0 \underline{1}') B(B'B)^{-1} B'(\underline{y} + \beta_0 \underline{1}) \right. \right. \\ &\quad \left. \left. + (\underline{y}' - \underline{e}') D(\underline{y} - \underline{e}) + b \right] \right\}. \end{aligned}$$

Note that

$$B(B'B)^{-1} B' = I - \underline{c}(\underline{c}'\underline{c})^{-1} \underline{c}'$$

where $\underline{c}' = (c_1, \dots, c_N)$ and I is the $N \times N$ identity matrix. Thus

$$\begin{aligned} \pi_a(y_{ns}, \beta_0, \theta, \sigma^2, \tau | I, y_s) &\propto \tau^{\frac{N-1}{2}} \left(\frac{1}{\sigma^2} \right)^{\frac{2N+a+1}{2}} \\ &\times \exp \left\{ -\frac{1}{2\sigma^2} \left[\tau(y' + \beta_0 \underline{1}') (I - \underline{c}(\underline{c}'\underline{c})^{-1}\underline{c})(y + \beta_0 \underline{1}) \right. \right. \\ &\quad \left. \left. + (y' - \underline{e}')D(y - \underline{e}) + b \right] \right\}. \end{aligned}$$

Now we may use the Gibbs sampler to draw a sample from $\pi_a(y_{ns}, \beta_0, \theta, \sigma^2, \tau | I, y_s)$ and the SIR algorithm to “convert” this sample to the one from $\pi(y_{ns}, \beta_0, \theta, \sigma^2, \tau | I, y_s)$. To perform the Gibbs sampler under $\pi_a(y_{ns}, \beta_0, \theta, \sigma^2, \tau | I, y_s)$, we need the conditional posterior density of each parameter given all the others. First we note that

$$y | \beta_0, \theta, \sigma^2, \tau \sim \text{Normal} \{ \Sigma D \underline{e} + (\Sigma D - I) \beta_0 \underline{1}, \sigma^2 \Sigma \}$$

where

$$\begin{aligned} \Sigma &= [(\tau I + D) - \tau \underline{c}(\underline{c}'\underline{c})^{-1}\underline{c}]^{-1} \\ &= (\tau I + D)^{-1} + \frac{1}{k} (\underline{c}^*)(\underline{c}^*)' \\ \underline{c}^* &= (\tau I + D)^{-1} \underline{c} \\ k &= \underline{c}' (\tau^{-1} \underline{c} - \underline{c}^*) > 0 \\ (\tau I + D)^{-1} &= \text{diag} \left[\frac{1}{\tau + (g'(a_1))^2}, \dots, \frac{1}{\tau + (g'(a_N))^2} \right]. \end{aligned}$$

Hence,

$$\begin{aligned} y_{ns} | y_s, \beta_0, \theta, \sigma^2, \tau &\sim \text{Normal} \left\{ \underline{e}_{ns} + \frac{\underline{c}'_s(y_s - \underline{e}_s)}{k_1} \underline{c}_{ns}^* + \tau \left[\frac{\tau^{-1} \underline{e}'_s \underline{c} - \underline{e}'_s \underline{c}^* + \underline{e}'_s \underline{c}_s^*}{\tau^{-1} \underline{c}'_s \underline{c} - \underline{c}'_s \underline{c}^* + \underline{c}'_s \underline{c}_s^*} \underline{c}_{ns}^* - \underline{e}_{ns}^* \right] \right. \\ &\quad \left. + \beta_0 \tau \left[\frac{\tau^{-1} \underline{1}'_s \underline{c} - \underline{1}'_s \underline{c}^* + \underline{1}'_s \underline{c}_s^*}{\tau^{-1} \underline{c}'_s \underline{c} - \underline{c}'_s \underline{c}^* + \underline{c}'_s \underline{c}_s^*} \underline{c}_{ns}^* - \underline{1}_{ns}^* \right], \right. \\ &\quad \left. \sigma^2 \left((\tau I_{ns} + D_{ns})^{-1} + \frac{1}{k_1} (\underline{c}_{ns}^*)(\underline{c}_{ns}^*)' \right) \right\} \end{aligned}$$

where

$$\begin{aligned}
k_1 &= k + \underset{\sim}{c}'_s \underset{\sim}{c}_s^* = \tau^{-1} \underset{\sim}{c}' \underset{\sim}{c} - \underset{\sim}{c}' \underset{\sim}{c}^* + \underset{\sim}{c}'_s \underset{\sim}{c}_s^* \\
\underset{\sim}{c}_s^* &= (\tau I_s + D_s)^{-1} \underset{\sim}{c}_s \\
\underset{\sim}{c}_{ns}^* &= (\tau I_{ns} + D_{ns})^{-1} \underset{\sim}{c}_{ns} \\
\underset{\sim}{e}_s^* &= (\tau I_s + D_s)^{-1} \underset{\sim}{e}_s \\
\underset{\sim}{e}_{ns}^* &= (\tau I_{ns} + D_{ns})^{-1} \underset{\sim}{e}_{ns} \\
\underset{\sim}{1}_s^* &= (\tau I_s + D_s)^{-1} \underset{\sim}{1}_s \\
\underset{\sim}{1}_{ns}^* &= (\tau I_{ns} + D_{ns})^{-1} \underset{\sim}{1}_{ns} \\
(\tau I_{ns} + D_{ns})^{-1} &= \text{diag} \left[\frac{1}{\tau + (g'(a_{n+1}))^2}, \dots, \frac{1}{\tau + (g'(a_N))^2} \right].
\end{aligned}$$

However, if we can assume

$$\theta_i = \alpha_0 + \alpha_1 x_i, \quad i = 1, \dots, N$$

where x_i are known and α_0, α_1 are the covariates of our interest, then the conditional marginal density

$$f(\underset{\sim}{y}_{ns} | \underset{\sim}{y}_s, \beta_0, \underset{\sim}{\alpha}, \sigma^2, \tau) = f(\underset{\sim}{y}_{ns} | \underset{\sim}{y}_s, \beta_0, \underset{\sim}{\theta}, \sigma^2, \tau),$$

where $\underset{\sim}{\alpha} = (\alpha_0, \alpha_1)'$. The conditional posterior densities of $\beta_0, \underset{\sim}{\alpha}, \sigma^2, \tau$ under $\pi_a(\underset{\sim}{y}_{ns}, \beta_0, \underset{\sim}{\alpha}, \sigma^2, \tau | \underset{\sim}{I}, \underset{\sim}{y}_s)$ are

$$\begin{aligned}
\beta_0 | \underset{\sim}{y}, \underset{\sim}{\alpha}, \sigma^2, \tau &\sim \text{Normal} \left(\frac{\bar{y}_w - \bar{y}}{1 - \frac{N}{\underset{\sim}{c}' \underset{\sim}{c}}}, \frac{\sigma^2}{\tau N (1 - \frac{N}{\underset{\sim}{c}' \underset{\sim}{c}})} \right), \\
\underset{\sim}{\alpha} | \underset{\sim}{y}, \beta_0, \sigma^2, \tau &\sim \text{Normal} \left((\underset{\sim}{x}' \underset{\sim}{x})^{-1} \underset{\sim}{x}' \underset{\sim}{g}, \sigma^2 (\underset{\sim}{x}' \underset{\sim}{x})^{-1} \right), \\
\sigma^{-2} | \underset{\sim}{y}, \beta_0, \underset{\sim}{\alpha}, \tau &\sim \text{Gamma} \left(\frac{2N + a - 1}{2}, \frac{\tau A + G + b}{2} \right), \\
\tau | \underset{\sim}{y}, \beta_0, \underset{\sim}{\alpha}, \sigma^2 &\sim \text{Gamma} \left(\frac{N + 1}{2}, \frac{A}{2\sigma^2} \right),
\end{aligned}$$

$$\begin{aligned}
\text{where } \bar{y}_w &= \frac{\sum_{i=1}^N c_i y_i}{\sum_{i=1}^N c_i^2}, \quad \bar{y} = \frac{\sum_{i=1}^N y_i}{N}, \quad \underset{\sim}{x}' = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_N \end{pmatrix} \\
\underset{\sim}{g} &= (g(a_1) + g'(a_1)(y_1 - a_1), \dots, g(a_N) + g'(a_N)(y_N - a_N))' \\
A &= \sum_{i=1}^N (\beta_0 + y_i)^2 - \frac{[\sum_{i=1}^N c_i (\beta_0 + y_i)]^2}{\sum_{i=1}^N c_i^2}, \quad G = \sum_{i=1}^N [g'(a_i)(y_i - e_i)]^2.
\end{aligned}$$

Finally, we note that we can draw $y_{ns}|\beta_0, \alpha, \sigma^2, \tau, y_s$ using the product rule. It is easy to show that

$$y_{n+k+1}|\underline{y}_{n+k}, \beta_0, \alpha, \sigma^2, \tau \sim \text{Normal} \left\{ e_{n+k+1} + \frac{\underline{c}'_{n+k}(\underline{y}_{n+k} - \underline{e}_{n+k})}{k_1^{(n+k)}} c_{n+k+1}^* + \right. \\ \left. \tau \left[\frac{\tau^{-1} \underline{e}' \underline{c} - \underline{e}' \underline{c}^* + \underline{e}'_{n+k} \underline{c}_{n+k}^*}{k_1^{(n+k)}} c_{n+k+1}^* - e_{n+k+1}^* \right] + \right. \\ \left. \beta_0 \tau \left[\frac{\tau^{-1} 1' \underline{c} - 1' \underline{c}^* + 1'_{n+k} \underline{c}_{n+k}^*}{k_1^{(n+k)}} c_{n+k+1}^* - 1^* \right], \right. \\ \left. \sigma^2 \left[1^* + \frac{(c_{n+k+1}^*)^2}{k_1^{(n+k)}} \right] \right\}$$

where, $c_{n+k+1}^* = \frac{c_{n+k+1}}{\tau + (g'(a_{n+k+1}))^2}$, $e_{n+k+1}^* = \frac{e_{n+k+1}}{\tau + (g'(a_{n+k+1}))^2}$, $1^* = \frac{1}{\tau + (g'(a_{n+k+1}))^2}$, and $k_1^{(n+k)} = \tau^{-1} \underline{c}' \underline{c} - \underline{e}' \underline{c}^* + \underline{e}'_{n+k} \underline{c}_{n+k}^*$. Thus by the product rule,

$$f(\underline{y}_{ns}|\underline{y}_s, \beta_0, \alpha, \sigma^2, \tau) = \prod_{k=n+1}^N f(y_k|\underline{y}_{(k)}, \beta_0, \alpha, \sigma^2, \tau)$$

where $\underline{y}_{(k)} = (y_1, \dots, y_{k-1})$, $k = n+1, \dots, N$.

The Gibbs sampler provides a sample $\Omega^{(h)}$, $h = 1, \dots, M$ from the joint posterior density, where $\Omega^{(h)} = (\underline{y}_{ns}, \beta_0, \alpha, \sigma^2, \tau)$. We perform the SIR algorithm by sub-sampling the $\Omega^{(h)}$ with weights

$$W_h = \frac{R(\Omega^{(h)})}{\sum_{h=1}^M R(\Omega^{(h)})}, \quad h = 1, \dots, M.$$

Then, we draw a sample from the discrete probability mass function $\{(\Omega^{(h)}, W_h), h = 1, \dots, M\}$ with replacement.

3.4 Discussion

Our model can be easily extended to a more generalized model. That is, in the model assumption, instead of assuming $\pi_i \propto \beta_0 + y_i + e_i$, $i = 1, \dots, N$, we may assume

$$\pi_i \propto \beta_0 + h(y_i) + e_i, \quad i = 1, \dots, N$$

where $h(y_i)$ is a function of y_i . Obviously, this is a more generalized assumption. To handle this more generalized model, we may transform y_i into z_i , where $z_i = h(y_i)$. Hence,

$$\pi_i \propto \beta_0 + z_i + e_i, \quad i = 1, \dots, N.$$

Now, instead of having $g(y_i)|\theta_i, \sigma^2 \stackrel{iid}{\sim} \text{Normal}(\theta_i, \sigma^2)$, $i = 1, \dots, N$, we will have a function g^* such that

$$g^*(z_i)|\theta_i^*, (\sigma^*)^2 \stackrel{iid}{\sim} \text{Normal}(\theta_i^*, (\sigma^*)^2), \quad i = 1, \dots, N.$$

Thus, we can use our model above over z_i to make inference. It will be interesting to see how well our model works in the situation where traditional regression method are not appropriate, i.e. we want to determine parameters β_0, β_1 in a linear relation

$$\nu_i = \beta_0 + \beta_1 y_i + e_i$$

given ν_i and y_i , $i = 1, \dots, N$. Where, $e_i \sim \text{Normal}(0, \sigma_e^2)$ and σ_e^2 is known, but $y_i \sim \text{Normal}(\mu, \sigma^2)$. This situation can not be well addressed by traditional regression method because it treats y_i as constants. However, in fact, the y_i , $i = 1, \dots, N$, are observed with unequal probabilities. Hence, the traditional regression method may give some biased estimates for β_0, β_1 . Our model may address the situation of this kind very well.

3.5 Data Simulation

For data simulation, we set $g(y) = \ln y$. Then we use both the ignorable and nonignorable models and follow the same procedure of data simulation as described in Chapter 2. Also, 1000 datasets are generated. For abbreviation, we only list the result with regard to the inference by the two models of population mean \bar{y} in the table as below.

Table 5: Comparison of the ignorable model with transformation (IGT) and the nonignorable model with transformation (NIGT) in case where small selection bias exists.

IG	NIG	RAVG	RSTD	<u>Coverage by 95% Interval</u>	
				IG	NIG
1.0107	1.0013	0.1052	0.0265	0.902	0.939

Table 6: Comparison of the ignorable model with transformation (IGT) and the nonignorable model with transformation (NIGT) in case where big selection bias exists.

IG	NIG	RAVG	RSTD	<u>Coverage by 95% Interval</u>	
				IG	NIG
1.0156	1.0020	0.1049	0.0313	0.885	0.918

Thus, the results are similar to those in the untransformed case.

Chapter 4

EXAMPLE ON NATURAL GAS PRODUCTION

4.1 Data description

In this chapter, we will apply our generalized model to a real problem to illustrate the potential of our model. Because the data are confidential, we have to hide its source; we will refer to the agency that provides the data as the *source*. We were told to consider a set of 492 natural gas companies as the population; the response was the average daily natural gas production for each of these 492 companies in 2002. We were provided with all 492 selection probabilities but only 31 companies were sampled using PPS sampling. The probabilities of selection are proportional to the average daily gas production in 2000. We were also told that the average daily production of these 492 companies is 518 million cubic feet.

In the table below we provide the names of the 31 sampled operators, together with the sample probabilities and the 2002 natural gas production in millions of cubic feet. We note that

- (a) The 2002 daily Natural Gas Production, y_i , by each of the 492 US operators.
- (b) The selection probability π_i for each of the 492 US operators in a sample survey for year 2002, which is proportional to the corresponding 2001 production.
- (c) In this case, we are informed of the 31 sampled operators out of the total 492 US operators. For convenience, we labeled the sampled operators as $i = 1, \dots, 31$ and nonsampled operators as $i = 32, \dots, 492$.

We want to make inference about the average daily gas production of these 492 companies using the biased sample of 31 companies. Specifically, we will consider point and interval estimation of the finite population mean as we obtained in Chapter 3.

Table 4.1: Average daily natural gas production (millions of cubic feet) for a sample of 31 operators in 2002 and the selection probabilities

Operator	Selection probability	Production
AMERADA HESS CORP	0.0592	455
ANADARKO PETROLEUM CORP	0.1721	1567
APACHE CORP	0.0630	1258
BP AMOCO PLC	0.5634	5090
BURLINGTON RESOURCES OIL, GAS CO	0.1700	2241
CABOT OIL, GAS CORP	0.0296	228
CHESAPEAKE OPERATING INC	0.0667	502
CHEVRON USA PRODUCTION CO	0.4182	3715
CONOCO INC	0.2972	2539
DEVON ENERGY CORP	0.2419	1525
EL PASO PRODUCTION CO	0.2177	1941
EQUITABLE PRODUCTION CO	0.0394	309
EXXON MOBIL PRODUCTION CO	0.4005	3712
FOREST OIL CORP	0.0339	295
HUNT OIL CO	0.0330	180
KAISER-FRANCIS OIL CO	0.0238	192
KERR MCGEE, OG CORP	0.1421	463
MARATHON OIL CO	0.1034	905
MERIT ENERGY CO	0.0182	337
NEWFIELD EXPLORATION CO	0.0715	449
OCEAN ENERGY INC	0.0554	505
PIONEER NATURAL RESOURCES USA	0.0468	395
QUESTAR EXPLORATION AND PRODUCTION	0.0345	183
SAMSON RESOURCES CO	0.0552	421
SHELL OIL CO	0.2876	2596
STONE ENERGY CORPORTATION	0.0355	273
THE HOUSTON EXPLORATION	0.0378	264
TOM BROWN INC	0.0200	202
TOTAL FINA ELF S.A	0.0446	222
UNOCAL CORP	0.1361	1251
WALTER OIL, GAS CORP	0.0316	294

NOTE: A PPS sample of 31 natural gas operators was taken from a popultaion of 492. The data are confidential, and we are not allowed to discuss the source of the data.

4.2 Simple Model Checking

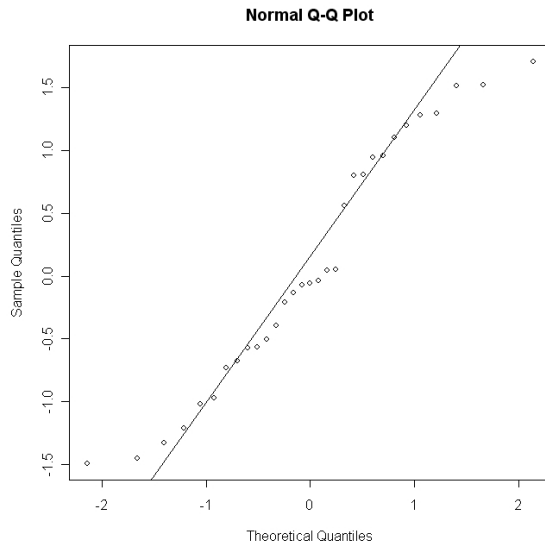
Before applying our model, we have to check that the assumptions of our model are satisfied by the real data.

(a) Normality of the Data.

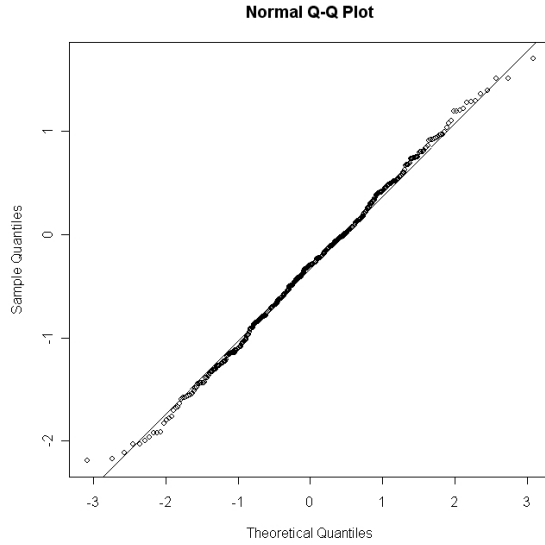
The sampled y_i , $i = 1, \dots, 31$ shows a dramatic right skewed pattern, which says they are not normal at all. Thus, first we need a transformation to bring the samples back to a normal distribution. We have tried many transformations with the Box-Cox family, but all of them fail the test of normality. After a time-consuming search we have found the transformation

$$g(y) = \ln(\sqrt{y} - \sqrt{120}).$$

As y in our data are all well above 120, this transformation will pose virtually no problem. In accordance with the variable names of our generalized model, we refer $g(y_i)$ as z_i here. The Shapiro-Wilk normality test for the transformed samples gives p-value 0.1180 and Kolmogorov-Smirnov test gives p-value 0.7022. Thus, for this transformation we do not reject the normality of the transformed samples. The Normal QQ plot also shows normality holds in the transformed sampled data z_i , $i = 1, \dots, 31$.



The source has generously provided an overall information of all 492 gas production. It checked the normality of the transformed 2002 production data of the 492 US operators under the same transformation. The Shapiro-Wilk normality test for the transformed data gives p-value 0.653 and Kolmogorov-Smirnov test gives p-value 0.8966. The QQ plot of the source also shows normality of the transformed data $z_i, i = 1, \dots, 492$.



(b) Linear Relation between $h(y_i)$ and π_i .

The source has also checked the relation between y_i and π_i according to our generalized model. It comes out that it is simply

$$h(y) = y.$$

The linear relation holds between y_i and π_i for all the 492 US operators. That is

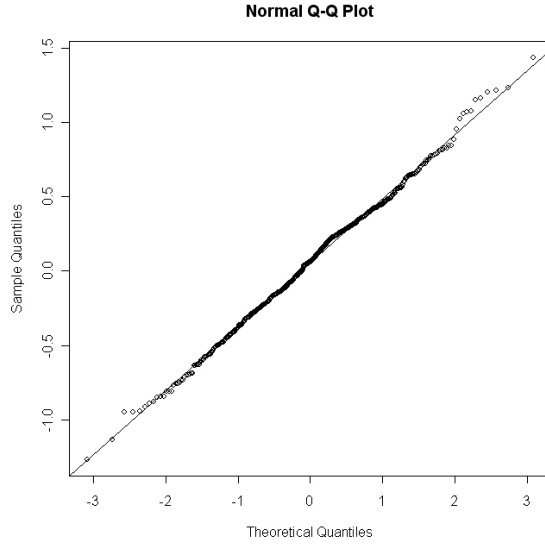
$$\pi_i = 0.0064681 + 0.0001092y_i + e_i$$

where, e_i is the residual, conforming to a normal distribution.

(c) Normality of the residues e_i

Finally, the source also checked the normality of the e_i . The Shapiro-Wilk normality test for the transformed samples gives p-value 0.7476 and Kolmogorov-Smirnov test

gives p-value 0.9973, which strongly suggest the normality of the residues. The Normal QQ plot below also shows that normality holds in the residuals z_i , $i = 1, \dots, 492$.



Now, since all the conditions of our generalized model are satisfied, we may apply it to the real data.

4.3 Computational Issues

We first regress y_i over c_i where, $c_i = \frac{492}{31}\pi_i$ for all $i = 1, \dots, 31$, it turns out to be $y_i = 576.64c_i - 52.93$, $i = 1, \dots, 31$; then we plug the given c_i , $i = 1, \dots, 495$ in the formula $a_i = 576.64c_i - 52.93$, $i = 1, \dots, 495$ to determine a_i , $i = 1, \dots, 495$. We will use the a_i as approximation to y_i in our model in the Taylor expansion of $g(y)$. For your reference, here we give the derivative of the transformation function as required in using the generalized model,

$$g'(y) = \frac{1}{2\sqrt{y}(\sqrt{y} - \sqrt{120})}.$$

4.4 Result

Table 7: Comparison of Ignorable Model and Nonignorable Model on Rescaled Real Data in inference of population mean \bar{z} , where $z = y/120$.

Statistical Fact of Rescaled Real Data ($z = y/120$)					
N	n	\bar{z}_s	$S_{z_s}^2$	\bar{z}	S_z^2
492	31	9.2766	111.1827	4.3166	18.0322
Model	$\hat{\bar{z}}$	$\sigma_{\hat{\bar{z}}}$	<u>95% interval for \bar{z}</u>		
NIG	5.1483	0.0190	5.1359	5.1766	
IG	15.3664	12.3576	6.2080	38.5749	

Note that the nonignorable model gives a point estimate of population mean \bar{z} of 5.1483. This is much closer to the true population mean 4.31655 (communicated to us by the source) than 15.3664 by ignorable model. The standard deviation for the estimate by nonignorable model is 0.0190 much smaller than that by ignorable model. However, the 95% interval for \bar{z} by both models doesn't contain the true value. This is because the selection bias is so large that the models can not correct the bias very well.

4.5 References

- Aires, N. (2003a), “Algorithms to find exact inclusion probabilities for conditional Poisson sampling,” *Methodology and Computing in Applied Probabilities*, to appear.
- Aires, N. (2000), “Comparisons between conditional Poisson sampling and Pareto pps sampling,” *Journal of Statistical Planning and Inference*, 82, 133-147.
- Aires, N. (2003b), “A guide to Fortran programs to calculate inclusion probabilities for sampling designs,” *Computational Statistics*, to appear.
- Burgos, C. G. B. and Nandram, B. (2003), “Bayesian Predictive Inference and Selection Bias,” *Journal of Research in Science and Engineering*, Vol. 1, No. 1, 1-10.
- Chambers, R., Dorfman, A., and Wang, S. (1998), “Limited Information Likelihood Analysis of Survey Data,” *Journal of the Royal Statistical Society, Series B*, 60, 397-411.
- Ghosh, D. and Vogt, A. (1998), “Rectification of sample size in Bernoulli and Poisson sampling,” *Proceedings of the American Statistical Association, Survey Research Methods Section*, pg., 1-6.
- Krieger, A. and Pfeffermann, D. (1992), “Maximum Likelihood Estimation From Complex Sample Surveys,” *Survey Methodology*, 18, 225-239.
- Malec, D., Davis, W. and Cao, X. (1999), “Model-based Small Area Estimates of Overweight Prevalence Using Sample Selection Adjustment,” *Statistics in Medicine*, 18, 3189-3200.
- Nandram, B. and Sedransk, J. (2004), “A Bayesian Predictive Inference for Finite Population Quantities Under Informative Sampling”, *Technical Report, Worcester Polytechnic Institute*.

- Patil, G. P. and Rao, C. R. (1978), “Weighted Distributions and Size-Biased Sampling With Applications to Wildlife Populations and Human Families,” *Biometrics*, 34, 179-189.
- Pfeffermann, D., Krieger, A. M., and Rinott, Y. (1998), “Parametric Distributions of Complex Survey Data under Informative Probability Sampling,” *Statistica Sinica*, 8, 1087-1114.
- Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., and Rasbash J. (1998), “Weighting for Unequal Selection Probabilities in Multilevel Models,” (with discussion), *Journal of the Royal Statistical Society, Series B*, 60, 23-40.
- Pfeffermann, D. and Sverchkov, M. (1999), “Parametric and Semi-parametric Estimation of Regression Models Fitted to Survey Data,” *Sankhya*, Series B, 61, 1-21.
- Sugden, R. A. and Smith, T. M. F. (1984), “Ignorable and Informative Designs in Survey Sampling,” *Biometrika*, 71, 495-506.
- You, Y. and Rao, J. N. K. (2003), “Pseudo Hierarchical Bayes Small Area Estimation Combining Unit Level Models and Survey Weights,” *Journal of Statistical Planning and Inference*, 111, 197-208.

Appendix A

MODEL WITH ALL INCLUSION PROBABILITIES UNKNOWN

A.1 Model Assumption

Let π_i , $i = 1, \dots, N$ denote the set of selection probabilities, $0 \leq \pi_i \leq 1$. Here we assume that the π_i are all unknown. Let y_i denote the corresponding response variable, $i = 1, \dots, N$. Then we assume that

$$\pi_i \propto \beta_0 + y_i + e_i, \quad i = 1, \dots, N$$

where e_i are errors. This assumption is reasonable for establishment survey. That is, the sample design is informative, and the π_i are “proportional” to the $\beta_0 + y_i$ with noise in the proportionality. Now any sample design must satisfy $\sum_{i=1}^N \pi_i = n$, the sample size. Thus,

$$\pi_i = \frac{n(\beta_0 + y_i + e_i)}{N(\beta_0 + \bar{y} + \bar{e})}, \quad i = 1, \dots, N$$

where $\bar{y} = \frac{\sum_{i=1}^N y_i}{N}$ and $\bar{e} = \frac{\sum_{i=1}^N e_i}{N}$.

Now letting $\nu_i = \beta_0 + y_i + e_i$, $i = 1, \dots, N$ where ν_i is a latent variable, we have

$$\pi_i = \frac{n\nu_i}{N\bar{\nu}}, \quad i = 1, \dots, N.$$

Note that because $\pi_i \geq 0$, the ν_i must all be non-negative or non-positive. We take $\nu_i \geq 0$, $i = 1, \dots, N$. Because $0 \leq \pi_i \leq 1$, it follows that,

$$0 \leq \nu_i \leq \frac{1}{n-1} \sum_{j \neq i} \nu_j.$$

Note that the selection indicator I_i are all observable, $I_i = 1$ if $i \in s$ and $I_i = 0$ if $i \notin s$. We also assume that the response variables

$$y_i | \mu, \sigma^2 \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2), \quad i = 1, \dots, N.$$

This is a standard assumption for a random sample drawn from the population also. However, because of the selection bias, this assumption fails for both the sampled individuals and the non-sampled individuals. Let \underline{y}_s denote the vector of sampled values, and \underline{y}_{ns} the vector of non-sampled values. Then the vector of all population values is $\underline{y} = (\underline{y}'_s, \underline{y}'_{ns})'$.

A.2 Main Results

Given $\beta_0, \mu, \sigma^2, \sigma_e^2$, (I_i, ν_i, y_i) are independent with joint density function, it follows that

$$\begin{aligned} P(\underline{I}, \underline{\nu}, \underline{y} | \beta_0, \mu, \sigma^2, \sigma_e^2) &= P(\underline{I} | \underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2, \sigma_e^2) \\ &\times P(\underline{\nu} | \underline{y}, \beta_0, \mu, \sigma^2, \sigma_e^2) P(\underline{y} | \beta_0, \mu, \sigma^2, \sigma_e^2) \\ &= \left\{ \prod_{i=1}^N \left(\frac{n\nu_i}{N\bar{\nu}} \right)^{I_i} \left(1 - \frac{n\nu_i}{N\bar{\nu}} \right)^{1-I_i} \right\} \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2} \\ &\times \frac{\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_e} e^{-\frac{1}{2\sigma_e^2}(\nu_i - \beta_0 - y_i)^2}}{\int_{V^{(n)}} \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_e} e^{-\frac{1}{2\sigma_e^2}(\nu_i - \beta_0 - y_i)^2} d\underline{\nu}}, \end{aligned}$$

where

$$\underline{\nu} = (\nu_1, \nu_2, \dots, \nu_N), \quad V^{(N)} = \{\underline{\nu} \mid 0 \leq \nu_i \leq \frac{1}{n-1} \sum_{j \neq i} \nu_j, \quad i = 1, \dots, N\}.$$

It is convenient to do a re-parameterization. We take $\rho = \frac{\sigma^2}{\sigma^2 + \sigma_e^2}$, so that $\sigma_e^2 = \frac{1-\rho}{\rho}\sigma^2$. Thus our new parameters are $\beta_0, \mu, \sigma^2, \rho$ and note that $0 < \rho < 1$. It is interesting that we now have ρ bounded. Before we had $\beta_0, \mu, \sigma^2, \sigma_e^2$ all unbounded. Thus we have

$$\begin{aligned} P(\underline{I}, \underline{\nu}, \underline{y} | \beta_0, \mu, \sigma^2, \rho) &= \left\{ \prod_{i=1}^N \left(\frac{n\nu_i}{N\bar{\nu}} \right)^{I_i} \left(1 - \frac{n\nu_i}{N\bar{\nu}} \right)^{1-I_i} \right\} \times \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2} \\ &\times \frac{\prod_{i=1}^N \left(\frac{\rho}{1-\rho} \right)^{\frac{1}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} \left(\frac{\rho}{1-\rho} \right) (\nu_i - \beta_0 - y_i)^2}}{\int_{V^{(n)}} \prod_{i=1}^N \left(\frac{\rho}{1-\rho} \right)^{\frac{1}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} \left(\frac{\rho}{1-\rho} \right) (\nu_i - \beta_0 - y_i)^2} d\underline{\nu}}. \end{aligned}$$

For a full Bayesian analysis, we take $\beta_0, \mu, \sigma^2, \rho$ to be independent with

$$P(\beta_0) = 1, \quad -\infty < \beta_0 < \infty,$$

$$P(\mu) = 1, \quad -\infty < \mu < \infty,$$

$$\sigma^{-2} \sim \text{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right), \quad a = .002, \quad b = .002,$$

and

$$\rho \sim \text{Uniform}(0, 1).$$

Thus, the joint prior density is

$$\pi(\beta_0, \mu, \sigma^2, \rho) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}}, \quad -\infty < \beta_0, \mu < \infty, \quad 0 < \rho < 1, \quad \sigma^2 > 0.$$

Note that this is a proper prior in σ^2 and ρ but not in β_0, μ . Thus, by Bayes' theorem the joint posterior density of $(\nu, y_{ns}, \beta_0, \mu, \sigma^2, \rho)$ is

$$\begin{aligned} P(\nu, y_{ns}, \beta_0, \mu, \sigma^2, \rho \mid I, y_s) &\propto P(I, \nu, y, \beta_0, \mu, \sigma^2, \rho) \\ &= P(I, \nu, y \mid \beta_0, \mu, \sigma^2, \rho) \pi(\beta_0, \mu, \sigma^2, \rho) \\ &= \left\{ \prod_{i=1}^N \left(\frac{n\nu_i}{N\bar{\nu}}\right)^{I_i} \left(1 - \frac{n\nu_i}{N\bar{\nu}}\right)^{1-I_i} \right\} \frac{\prod_{i=1}^N \left(\frac{\rho}{1-\rho}\right)^{\frac{1}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\left(\frac{\rho}{1-\rho}\right)(\nu_i - \beta_0 - y_i)^2}}{\int_{V(n)} \prod_{i=1}^N \left(\frac{\rho}{1-\rho}\right)^{\frac{1}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\left(\frac{\rho}{1-\rho}\right)(\nu_i - \beta_0 - y_i)^2} d\nu} \\ &\times \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2} \left(\frac{1}{\sigma^2}\right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}}. \end{aligned}$$

It is convenient to transform ρ to $\tau = \frac{\rho}{1-\rho}$ keeping $\nu, y_{ns}, \beta_0, \mu, \sigma^2$ untransformed. Thus,

$$\begin{aligned} P(\nu, y_{ns}, \beta_0, \mu, \sigma^2, \tau \mid I, y_s) &\propto \frac{\prod_{i=1}^N \left(\frac{\tau}{2\pi\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\tau}{2\sigma^2}(\nu_i - \beta_0 - y_i)^2}}{\int_{V(n)} \prod_{i=1}^N \left(\frac{\tau}{2\pi\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\tau}{2\sigma^2}(\nu_i - \beta_0 - y_i)^2} d\nu} \\ &\times \left\{ \prod_{i=1}^N \left(\frac{n\nu_i}{N\bar{\nu}}\right)^{I_i} \left(1 - \frac{n\nu_i}{N\bar{\nu}}\right)^{1-I_i} \right\} \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2} \left(\frac{1}{\sigma^2}\right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}} \left(\frac{1}{1+\tau}\right)^2, \end{aligned}$$

where $\left(\frac{1}{1+\tau}\right)^2$ is the Jacobian of the transformation.

A.3 Computation

To perform the computation, we write

$$P(\underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2, \sigma_e^2, I) = R(\underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2, \sigma_e^2, I) P_a(\underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2, \sigma_e^2)$$

where

$$\begin{aligned} P_a(\underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2, \sigma_e^2) &= \prod_{i=1}^N \left(\left(\frac{\tau}{2\pi\sigma^2} \right)^{\frac{1}{2}} e^{-\frac{\tau}{2\sigma^2}(\nu_i - \beta_0 - y_i)^2} \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2} \right) \\ &\times \left(\frac{1}{\sigma^2} \right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma^2}} \left(\frac{1}{\sigma_e^2} \right)^{\frac{a}{2}+1} e^{-\frac{b}{2\sigma_e^2}}, \\ R(\underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2, \sigma_e^2, I) &= \frac{\left\{ \prod_{i=1}^N \left(\frac{n\nu_i}{N\bar{\nu}} \right)^{I_i} \left(1 - \frac{n\nu_i}{N\bar{\nu}} \right)^{1-I_i} \right\} \left(\frac{\tau}{1+\tau} \right)^2}{\int_{V^{(n)}} \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_e} e^{-\frac{1}{2\sigma_e^2}(\nu_i - \beta_0 - y_i)^2} d\underline{\nu}}. \end{aligned}$$

Thus, we may use Gibbs sampler to generate sample of $(\underline{\nu}, \underline{y}_{ns}, \beta_0, \mu, \sigma^2, \tau)$ from $P_a(\underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2, \tau)$. To do this, we need conditional posterior density of each parameter given all the others. The conditional posterior densities of $\underline{\nu}, \underline{y}_{ns}, \beta_0, \mu, \sigma^2, \tau$ under $P_a(\underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2, \tau)$ are

$$\begin{aligned} \underline{\nu} | \underline{y}, \beta_0, \mu, \sigma^2, \tau &\sim \text{Normal} \left(\beta_0 \underline{1} + \underline{y}, \frac{\sigma^2}{\tau} I \right), \\ \underline{y}_{ns} | \underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2, \tau &\sim \text{Normal} \left(\frac{\tau}{1+\tau} (\underline{\nu}_{ns} - \beta_0 \underline{1}_{ns}) + \frac{\mu}{1+\tau} \underline{1}_{ns}, \frac{\sigma^2}{1+\tau} I \right), \\ \beta_0 | \underline{\nu}, \underline{y}, \mu, \sigma^2, \tau &\sim \text{Normal} \left(\bar{\nu} - \bar{y}, \frac{\sigma^2}{\tau N} \right), \\ \mu | \underline{\nu}, \underline{y}, \beta_0, \sigma^2, \tau &\sim \text{Normal} \left(\bar{y}, \frac{\sigma^2}{N} \right), \end{aligned}$$

and, letting $A = \sum_{i=1}^N (\nu_i - \beta_0 - y_i)^2$,

$$\begin{aligned} \sigma^{-2} | \underline{\nu}, \underline{y}, \beta_0, \mu, \tau &\sim \text{Gamma} \left(\frac{2N+a}{2}, \frac{\tau A + \sum_{i=1}^N (y_i - \mu)^2 + b}{2} \right), \\ \tau | \underline{\nu}, \underline{y}, \beta_0, \mu, \sigma^2 &\sim \text{Gamma} \left(\frac{N+2}{2}, \frac{A}{2\sigma^2} \right). \end{aligned}$$

The Gibbs sampler provides a sample $\Omega^{(h)}$, $h = 1, \dots, M$ from the joint posterior density, where $\Omega^{(h)} = (\underline{y}_{ns}, \beta_0, \mu, \sigma^2, \tau)$. We perform the SIR algorithm by sub-sampling the $\Omega^{(h)}$

with weights.

$$W_h = \frac{R(\Omega^{(h)})}{\sum_{h=1}^M R(\Omega^{(h)})}, \quad h = 1, \dots, M.$$

Then, we draw a sample from the discrete probability mass function $\{(\Omega^{(h)}, W_h), h = 1, \dots, M\}$ with replacement.

Appendix B

MODIFIED POISSON SAMPLING

B.1 Background

Suppose we have a population of N units, namely A_1, \dots, A_N . Each unit A_i has an inclusion probability π_i , $i = 1, \dots, N$. We intend to draw random sample of n units from the population. (Here, we consider cases where $n \geq 2$.) We assume $\sum_{i=1}^N \pi_i = n$. We hope that the sample obtained can numerically represent those inclusion probabilities. To solve this problem, we introduce a sampling method, referred as “Modified Poisson Sampling Method”.

Let X_i be the indicator of unit A_i being chosen, $i = 1, \dots, N$. Specifically, for each X_i , we draw a number u_i from Uniform(0,1). If $u_i \leq \pi_i$, then $X_i = 1$, otherwise $X_i = 0$. Obviously, in this way each A_i has the probability π_i to be chosen, $i = 1, \dots, N$ and for any $i \neq j$, A_i to be chosen is independent of A_j to be chosen. Running from $i = 1$ to $i = N$, we can obtain a sample in which each unit A_i has inclusion probability π_i . But the problem for this sampling method is that we may NOT get a sample of size n . To ensure we obtain sample of size n , we may drop off a sample if its size not equal to n . However, this will result in conditional probability and the final inclusion probability for each A_i may NOT be π_i .

This problem has been considered by others. Ghosh and Vogt (1998) provide a solution that is closest in spirit to ours. But their Poisson sampling is for small populations. Aires (2000) has looked at comparisons between Poisson sampling and Pareto PPS sampling. Some computational algorithms are available for Poisson sampling (see Aires 2003 a, b). Again these algorithms are limited to draws from small populations. Our methodology covers any population.

B.2 Existence of selection probabilities

As we retain samples only of size n and want inclusion probability for each unit A_i still being π_i , this actually is equivalent to require

$$P(X_i = 1 | \sum_{j=1}^N X_j = n) = \pi_i$$

where,

$$P(X_i = 1 | \sum_{j=1}^N X_j = n) = \frac{P(X_i = 1)P(\sum_{j \neq i} X_j = n-1)}{P(\sum_{j=1}^N X_j = n)}, \quad i = 1, \dots, N. \quad (\text{B.1})$$

Note that, if now we set the prior $P(X_i = 1) = \pi_i$, then the relation will not be automatically guaranteed. So we need to figure out the selection probabilities $P(X_i = 1)$ properly such that $P(X_i | \sum_{j=1}^N X_j = n) = \pi_i$ are satisfied.

Let $P(X_i = 1) = P_i$, $i = 1, \dots, N$, then

$$\begin{aligned} P(\sum_{j=1}^N X_j = n) &= \sum_{1 \leq j_1 < \dots < j_n \leq N} P(X_{j_1} = 1, \dots, X_{j_n} = 1, X_{j_{n+1}} = 0, \dots, X_{j_N} = 0) \\ &= \sum_{1 \leq j_1 < \dots < j_n \leq N} \cdot \left(\prod_{k=1}^n P_{j_k} \right) \left(\prod_{k=n+1}^N (1 - P_{j_k}) \right) \end{aligned}$$

For abbreviation, we denote

$$\prod_{k=1}^n P_{j_k} = \prod_{j \in S(n)} P_j, \quad \prod_{k=1}^n (1 - P_{j_k}) = \prod_{j \notin S(n)} (1 - P_j)$$

where $S(n) = \{(j_1, \dots, j_n) | 1 \leq j_1 < \dots < j_n \leq N\}$.

So,

$$\begin{aligned} P(\sum_{j=1}^N X_j = n) &= \sum_{S(n)} \left(\prod_{j \in S(n)} P_j \prod_{j \notin S(n)} (1 - P_j) \right) \\ &= \left(\prod_{j=1}^N (1 - P_j) \right) \left(\sum_{S(n)} \prod_{j \in S(n)} \left(\frac{P_j}{1 - P_j} \right) \right). \end{aligned}$$

Similarly,

$$\begin{aligned} P(\sum_{j \neq i} X_j = n-1) &= \sum_{S_i(n-1)} \left(\prod_{j \in S_i(n-1)} P_j \prod_{j \notin S_i(n-1)} (1-P_j) \right) \\ &= \left(\prod_{j \neq i} (1-P_j) \right) \left(\sum_{S_i(n-1)} \prod_{j \in S_i(n-1)} \left(\frac{P_j}{1-P_j} \right) \right) \end{aligned}$$

where $S_i(n-1) = \{(j_1, \dots, j_{n-1}) \mid 1 \leq j_1 < \dots < j_{n-1} \leq N, j_1, \dots, j_{n-1} \neq i\}$.

Then, (??) becomes

$$\frac{P_i}{1-P_i} \left(\sum_{S_i(n-1)} \prod_{j \in S_i(n-1)} \frac{P_j}{1-P_j} \right) / \left(\sum_{S_i(n)} \prod_{j \in S_i(n)} \frac{P_j}{1-P_j} \right) = \pi_i ,$$

$i = 1, \dots, N$.

Let the odds $\frac{P_i}{1-P_i} = q_i$, $i = 1, \dots, N$. Then,

$$q_i \left(\sum_{S_i(n-1)} \prod_{j \in S_i(n-1)} q_j \right) / \left(\sum_{S_i(n)} \prod_{j \in S_i(n)} q_j \right) = \pi_i ,$$

$i = 1, \dots, N$.

Now, to figure out the selection probabilities $(P_1, \dots, P_N)'$ is equivalent to solve the N -equation group of $(q_1, \dots, q_N)'$ as above. Obviously, if the solution to this equation group exists, then the solution is not unique. That is, if $\underline{q} = (q_1, \dots, q_N)'$ is a solution, then for any $k \in R^+$, $k\underline{q}$ is also a solution to the equation group. However, first of all, we need to ensure the solution to this equation group exists.

For convenience, let us define

$$\underline{q} = (q_1, \dots, q_N)',$$

$$\underline{f}(\underline{q}) = (f_1(\underline{q}), \dots, f_N(\underline{q}))' \tag{B.2}$$

where,

$$f_i(\underline{q}) = \pi_i \left(\sum_{S_i(n)} \prod_{j \in S_i(n)} q_j \right) / \left(\sum_{S_i(n-1)} \prod_{j \in S_i(n-1)} q_j \right), \tag{B.3}$$

$i = 1, \dots, N$.

Then (??) becomes

$$f(q) = q. \quad (\text{B.4})$$

Before we prove the existence of the solution to the equation (??), like $S(n)$ and $S_i(n-1)$, we define

$$S(n-1) = \{(j_1, \dots, j_{n-1}) \mid 1 \leq j_1 < \dots < j_{n-1} \leq N\}$$

and

$$S_i(n) = \{(j_1, \dots, j_n) \mid 1 \leq j_1 < \dots < j_n \leq N, j_1, \dots, j_n \neq i\}.$$

Similarly,

$$S_{ij}(n) = \{(j_1, \dots, j_n) \mid 1 \leq j_1 < \dots < j_n \leq N, j_1, \dots, j_n \neq i \text{ or } j\}$$

$$S_{ij}(n-1) = \{(j_1, \dots, j_{n-1}) \mid 1 \leq j_1 < \dots < j_{n-1} \leq N, j_1, \dots, j_{n-1} \neq i \text{ or } j\}.$$

For convenience, we let

$$\begin{aligned} G(n, N) &= \sum_{S(n)} \prod_{j \in S(n)} q_j, \\ G(n-1, N) &= \sum_{S(n-1)} \prod_{j \in S(n-1)} q_j, \\ G_i(n, N-1) &= \sum_{S_i(n)} \prod_{j \in S_i(n)} q_j, \\ G_i(n-1, N-1) &= \sum_{S_i(n-1)} \prod_{j \in S_i(n-1)} q_j, \\ G_{ij}(n, N-2) &= \sum_{S_{ij}(n)} \prod_{k \in S_{ij}(n)} q_k, \\ G_{ij}(n-1, N-2) &= \sum_{S_{ij}(n-1)} \prod_{k \in S_{ij}(n-1)} q_k. \end{aligned}$$

Obviously, in light of this notation,

$$\sum_{i=1}^N q_i = G(1, N), \quad \prod_{i=1}^N q_i = G(N, N).$$

To prove the existence of solution for equation (??), we need some lemmas.

Lemma 1. $G(n, N) = q_i G_i(n-1, N-1) + G_i(n, N-1).$

Proof: By definition,

$$\begin{aligned}
 G(n, N) &= \sum_{S(n)} \prod_{j \in S(n)} q_j \\
 &= \sum_{S_i(n-1)} \left(q_i \prod_{j \in S_i(n-1)} q_j \right) + \sum_{S_i(n)} \prod_{j \in S_i(n)} q_j \\
 &= q_i \sum_{S_i(n-1)} \prod_{j \in S_i(n-1)} q_j + \sum_{S_i(n)} \prod_{j \in S_i(n)} q_j \\
 &= q_i G_i(n-1, N-1) + G_i(n, N-1).
 \end{aligned}$$

Lemma 2. Let $q_i \geq 0, i = 1, \dots, N$. Then

$$q_j \geq q_i \Leftrightarrow G_i(n, N-1) \geq G_j(n, N-1)$$

$$q_j \geq q_i \Leftrightarrow q_j G_j(n-1, N-1) \geq q_i G_i(n-1, N-1).$$

Proof: By Lemma 1,

$$G_i(n, N-1) = q_j G_{ij}(n-1, N-2) + G_{ij}(n, N-2)$$

$$\text{and } G_j(n, N-1) = q_i G_{ij}(n-1, N-2) + G_{ij}(n, N-2).$$

So that $G_i(n, N-1) - G_j(n, N-1) = (q_j - q_i) G_{ij}(n-1, N-2)$. Hence,

$$q_j \geq q_i \Leftrightarrow G_i(n, N-1) \geq G_j(n, N-1).$$

Also by Lemma 1,

$$G(n, N) = q_i G_i(n-1, N-1) + G_i(n, N-1)$$

$$\text{and } G(n, N) = q_j G_j(n-1, N-1) + G_j(n, N-1).$$

So that $G_i(n, N) - G_j(n, N) = q_j G_j(n-1, N-1) - q_i G_i(n-1, N-1)$. Hence,

$$G_i(n, N) \geq G_j(n, N) \Leftrightarrow q_j G_j(n-1, N-1) \geq q_i G_i(n-1, N-1).$$

$$\text{So } q_j \geq q_i \Leftrightarrow q_j G_j(n-1, N-1) \geq q_i G_i(n-1, N-1).$$

Lemma 3. Let $0 < \pi_i \leq \pi_j$, $0 < \frac{q_i}{\pi_i} \leq \frac{q_j}{\pi_j}$, $\forall 1 \leq i < j \leq N$, then

$$\frac{q_i G_i(n-1, N-1)}{\pi_i} \leq \frac{q_j G_j(n-1, N-1)}{\pi_j}.$$

Proof: For $\forall i < j$, let

$$A = \{q_1^*, \dots, q_N^*\}, \text{ where } q_k^* = \begin{cases} q_k & \text{if } k \neq i, 1 \leq k \leq N \\ \frac{q_i}{\pi_i} & \text{if } k = i \end{cases}$$

$$B = \{q_1^{**}, \dots, q_N^{**}\}, \text{ where } q_k^{**} = \begin{cases} q_k & \text{if } k \neq j, 1 \leq k \leq N \\ \frac{q_j}{\pi_j} & \text{if } k = j \end{cases},$$

$$G^*(n, N) = \sum_{S(n)} \prod_{k \in S(n)} q_k^*,$$

$$G^{**}(n, N) = \sum_{S(n)} \prod_{k \in S(n)} q_k^{**}.$$

Since $\frac{q_i}{\pi_i} \leq \frac{q_j}{\pi_j}$, then

$$G^*(n, N) \leq G^{**}(n, N).$$

By lemma 1,

$$G^*(n, N) = q_i^* G_i^*(n-1, N-1) + G_i^*(n, N-1) = \frac{q_i}{\pi_i} G_i(n-1, N-1) + G_i(n, N-1)$$

and

$$G^{**}(n, N) = q_j^{**} G_j^{**}(n-1, N-1) + G_j^{**}(n, N-1) = \frac{q_j}{\pi_j} G_j(n-1, N-1) + G_j(n, N-1).$$

Because $0 < \frac{q_i}{\pi_i} \leq \frac{q_j}{\pi_j}$, and $0 < \pi_i \leq \pi_j$, so $q_i \leq q_j$. Hence, by Lemma 2

$$G_j(n, N-1) \leq G_i(n, N-1).$$

Note that

$$\frac{q_i}{\pi_i} G_i(n-1, N-1) - \frac{q_j}{\pi_j} G_j(n-1, N-1) = [G^*(n, N) - G^{**}(n, N)] + [G_j(n, N-1) - G_i(n, N-1)].$$

So

$$\frac{q_i}{\pi_i}G_i(n-1, N-1) - \frac{q_j}{\pi_j}G_j(n-1, N-1) \leq 0.$$

That is,

$$\frac{q_i G_i(n-1, N-1)}{\pi_i} \leq \frac{q_j G_j(n-1, N-1)}{\pi_j}.$$

As a summary, by Lemma 2 and Lemma 3, we have that if $\frac{q_i}{q_j} \leq \frac{\pi_i}{\pi_j} \leq 1$ (of course we require π_k and q_k be positive numbers, $k = 1, \dots, N$), then

$$\frac{q_i}{q_j} \leq \frac{q_i G_i(n-1, N-1)}{q_j G_j(n-1, N-1)} \leq \frac{\pi_i}{\pi_j} \leq 1 \leq \frac{G_i(n-1, N-1)}{G_j(n-1, N-1)}$$

Lemma 4. $\sum_{i=1}^N q_i G_i(n-1, N-1) = nG(n, N).$

Proof: We note the following three points;

- a. Any term in $\sum_{i=1}^N q_i G_i(n-1, N-1)$ must be a term in the expansion of $G(n, N)$.
- b. Any term in the expansion of $G(n, N)$ must be a term in $\sum_{i=1}^N q_i G_i(n-1, N-1)$, and each term in the expansion of $G(n, N)$ has n “images” in $\sum_{i=1}^N q_i G_i(n-1, N-1)$.
- c. $G(n, N)$ has $N \binom{N-1}{n-1}$ terms. $\sum_{i=1}^N q_i G_i(n-1, N-1)$ has $\binom{N}{n}$ terms. And $N \binom{N-1}{n-1} = n \binom{N}{n}$.

By a, b, c , we know $\sum_{i=1}^N q_i G_i(n-1, N-1) = nG(n, N).$

Now we can proceed to prove our main result. Without loss of generality, we may assume $0 < \pi_1 \leq \dots \leq \pi_N < 1$.

Theorem: Equation $\underline{f}(q) = q$ has solution on the close domain D , where $\underline{f}(q)$ is as defined in (??) and (??),

$$D = \left\{ (q_1, \dots, q_N) \left| a \leq \frac{q_1}{\pi_1} \leq \dots \leq \frac{q_N}{\pi_N} \leq b \right. \right\},$$

a, b are any constants satisfying $0 < a < b$.

Proof:

For abbreviation, in our following proof we denote $f_i(q) = f_i$.

1. Because $q \in D$, so

$$0 < q_i \leq q_j, \quad \forall 1 \leq i < j \leq N.$$

By Lemma 2,

$$G_i(n, N-1) \geq G_j(n, N-1).$$

Thus

$$\frac{f_i}{\pi_i} = \frac{G(n, N)}{G_i(n-1, N-1)} \leq \frac{G(n, N)}{G_j(n-1, N-1)} = \frac{f_j}{\pi_j},$$

That is,

$$\frac{f_i}{\pi_i} \leq \frac{f_j}{\pi_j}, \quad \forall 1 \leq i < j \leq N.$$

2. Because

$$\begin{aligned} f_1 &= \pi_1 \frac{G(n, N)}{G_1(n-1, N-1)}, \\ f_1 &= \sum_{j=1}^N \pi_1 \frac{q_j G_j(n-1, N-1)}{n G_1(n-1, N-1)}, \quad (\text{lemma 4}) \\ &\geq \sum_{j=1}^N \pi_j \frac{q_1 G_1(n-1, N-1)}{n G_1(n-1, N-1)}, \quad (\text{lemma 3}) \\ &= q_1. \end{aligned}$$

Hence,

$$\frac{f_1}{\pi_1} \geq \frac{q_1}{\pi_1} \geq a$$

Similarly,

$$\begin{aligned} f_N &= \sum_{j=1}^N \pi_N \frac{q_j G_j(n-1, N-1)}{n G_N(n-1, N-1)} \quad (\text{lemma 4}) \\ &\leq \sum_{j=1}^N \pi_j \frac{q_N G_N(n-1, N-1)}{n G_N(n-1, N-1)} \quad (\text{lemma 3}) \\ &= q_N. \end{aligned}$$

Hence,

$$\frac{f_N}{\pi_N} \leq \frac{q_N}{\pi_N} \leq b.$$

By 1 and 2, we have that, if $q \in D$, then $f_i(q) \in D$.

Before we apply the fixed point theorem, we need to check the continuity of $\underline{f}(q)$ on D .

Obviously, continuity of $\underline{f}(q)$ is guaranteed on D . Hence, by the fixed point theorem, there is a solution to the equation $\underline{f}(q) = q$ on the close domain D .

Last, we want to point out that, if π_i are not all equal, $i = 1, \dots, N$, then $\frac{q_1}{q_N} \leq \frac{\pi_1}{\pi_N} < 1$.

Hence

$$a \leq \frac{q_1}{\pi_1} < \frac{f_1}{\pi_1} < \frac{f_N}{\pi_N} < \frac{q_N}{\pi_N} \leq b$$

which implies $q = (q_1, \dots, q_N)'$ can not be the solution, where $\frac{q_1}{\pi_1} = \dots = \frac{q_N}{\pi_N}$. Further, note that the above inequality also implies

$$\frac{q_1}{q_1^{(0)}} < \frac{f_1}{q_1^{(0)}} \quad \text{and} \quad \frac{f_N}{q_N^{(0)}} < \frac{q_N}{q_N^{(0)}}$$

where $q_1^{(0)} = \frac{\pi_1}{1-\pi_1}$, $q_1^{(0)} = \frac{\pi_N}{1-\pi_N}$. Hence $q = (q_1, \dots, q_N)'$ can not be the solution either, where $\frac{q_1}{q_1^{(0)}} = \dots = \frac{q_N}{q_N^{(0)}}$. This fact answers the question in the beginning of this section about why in general the inclusion probabilities can not be used as selection probabilities for the Poisson Sampling.

B.3 Computation of selection probabilities

We use iteration to seek the solution $(q_1, \dots, q_N)'$. That is,

$$q^{(n+1)} = \underline{f}(q^{(n)})$$

where $q^{(n)}$ is estimation for a solution q by the n^{th} iteration. Because $\underline{f}(q)$ is continuous on D , so

$$\lim_{n \rightarrow \infty} q^{(n+1)} = \lim_{n \rightarrow \infty} \underline{f}(q^{(n)})$$

which implies,

$$q = \underline{f}(q).$$

For initial value, we set

$$\underline{q}^{(0)} = (\frac{\pi_1}{1 - \pi_1}, \dots, \frac{\pi_N}{1 - \pi_N})'.$$

It is easy to verify that $\underline{q}^{(0)} \in D$. Thus the iteration will produce a solution in D . Note that as we mentioned before, the solution is not unique. But any of them shall be able to give us answer to the selection probabilities P_i . That is,

$$P_i = \frac{q_i}{1 + q_i} \ , \ i = 1, \dots, N$$

where $(q_1, \dots, q_N)'$ is the solution.

For the calculation of $G(n, N)$ in $\underline{f}(\underline{q})$, we define a two-dimension function as follows,

$$G(j, k) = \begin{cases} 0 & \text{if } j > k \\ 1 & \text{if } j = 0 \\ q_k G(j - 1, k - 1) + G(j, k - 1) & \text{else} \end{cases}$$

where j, k are non-negative integers and all the q_k are given, $k = 1, 2, \dots, N$. Obviously, computation of $G(n, N)$ through $G(j, k)$ is supported by Lemma 1. By a dual iteration on j and k , we can also easily obtain $G_i(n, N)$. Thus $\underline{f}(\underline{q})$ can be calculated easily even when both the sample size n and population size N are pretty large.